Feasibility study on the use of Trigger-Object Level Analysis in the Search for the Standard Model Higgs boson produced by vector-boson fusion and decaying to bottom quarks with the ATLAS detector.

Andrew J Strange

School of Physics and Astronomy

2018

A dissertation submitted to the University of Manchester
for the degree of Master of Science by Research
in the Faculty of Engineering and Physical Sciences

Contents

WORD COUNT: ???

List of Figures

List of Tables

ABSTRACT

Write the abstract here.

Declaration

No portion of the work referred to in the dissertation has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

The author of this dissertation (including any appendices and/or schedules to this dissertation) owns certain copyright or related rights in it (the "Copyright") and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this dissertation, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the dissertation, for example graphs and tables ("Reproductions"), which may be described in this dissertation, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this dissertation, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy, in any relevant Dissertation restriction declarations deposited in the University Library, The University Library's regulations and in The University''s policy on Presentation of Dissertations

Acknowledgements

These are the acknowledgements.

CHAPTER 1

THEORY

1.1 Standard Model

The Standard Model (SM) of particle physics is a collection of several theories which provide the most accurate theoretical framework for describing all known components of matter and their interactions to date. The model describes three fundamental forces, each mediated by an integer spin particle called a *gauge boson*, that control interactions between the spin-$\frac{1}{2}$ *quarks* and *leptons* that make up matter. The mathematical structure is based on the symmetry group $SU(3)_c \times SU(2)_L \times U(1)_Y$ and is required to be gauge-invariant. The SM does not include gravity; gravity cannot be written in the Quantum Field Theories that describe the Standard Model, and gravitational interactions are significantly weaker than the other fundamental forces (Table **??**). As a result, gravitational interactions so are neglected hereafter.

1.1.1 Fermions

The full set of spin-$\frac{1}{2}$ *fermions*, described in Tables **??** and **??**, are the quark and lepton families, which each have three generations. For each distinct particle there is a paired *anti-particle* which is identical aside from opposite charge and *handedness*. The handedness or helicity of a particle refers to the projection of the angular momentum of the particle along the direction of the particle momentum. For a spin $\frac{1}{2}$ particle, the angular momentum component can be aligned along the direction of motion (*positive* or *right-handed* alignment) or opposed to it (*negative* or *left-handed* alignment)).

Most matter consists of the observable first generation of the up and down quarks and the electron which make up protons and neutrons, along with the unobservable electron neutrino. Both the leptons and the quarks obey Fermi-Dirac statistics. Quarks experience all three fundamental forces, charged leptons interacting via the electromagnetic and weak interactions and neutrinos experiencing only the weak interaction. Neutrinos have a special individual feature in that only left-handed neutrinos and right-handed anti-neutrinos have been observed. This asymmetry violates invariance under the *charge* (C) quantum operation and under the *parity* (P) operation individually, but does preserve CP invariance [**?**].

Table 1.1: Spin-$\frac{1}{2}$ fermions: quarks q [5]. The top quark mass is taken from direct measurements.

Generation	Flavour	Charge / e	Mass / GeV
1	Up u	+2/3	$0.0022\ ^{+0.0006}_{-0.0004}$
	Down d	-1/3	$0.0047\ ^{+0.0005}_{-0.0004}$
2	Charm c	+2/3	1.28 ± 0.03
	Strange s	-1/3	$0.096\ ^{+0.008}_{-0.004}$
3	Top t	+2/3	173.1 ± 0.6
	Bottom b	-1/3	$4.18\ ^{+0.04}_{-0.03}$

Table 1.2: Spin-$\frac{1}{2}$ fermions: leptons l [5]

Generation	Flavour	Charge / e	Mass / MeV
1	Electron e	-1	$0.5109989461 \pm 0.0000000031$
	Electron Neutrino ν_e	0	$< 2 \times 10^{-6}$
2	Muon μ	-1	$105.6583745 \pm 0.0000024$
	Muon Neutrino ν_μ	0	$< 2 \times 10^{-6}$
3	Tau τ	-1	1776.86 ± 0.12
	Tau Neutrino ν_τ	0	$< 2 \times 10^{-6}$

Quarks are always confined into colour singlet *hadrons* bound by the strong interaction, which are either *baryons* (qqq) like the *proton* (uud) and *neutron* (ddu), or *mesons* ($q\bar{q}$) like the positive *pion* ($u\bar{d}$).

1.1.2 Forces

All forces arise due the exchange of unobservable virtual particles, gauge bosons, which obey Bose-Einstein statistics. The three fundamental particle interactive forces for the SM are named the strong, weak and electromagnetic interactions, and are mediated by *gluons*, *weak bosons* and *photons* respectively. The gauge bosons are described in more detail in Table **??**.

Table 1.3: Spin-1 gauge bosons. The strength of the interaction is typically stated in terms of α, a dimensionless constant proportional to the matrix element for the virtual particle exchange for each interaction. For the The weak interaction is intrinsically stronger than the EM interaction, but the mass of the weak bosons limits the range to extremely short distances. The strength of gravity is $\sim 10^{-39}$ hence it is neglected. [5]

Interaction	Particle	Charge / e	Mass / GeV	Strength (α)
Strong	Gluon g	0	0	~ 1
Weak (Charged Current)	W^+	1	80.385 ± 0.015	
	W^-	-1	80.385 ± 0.015	10^{-6}
Weak (Neutral Current)	Z	0	91.1876 ± 0.0021	
Electromagnetic (EM)	Photon γ	$< 1 \times 10^{-35}$	$< 1 \times 10^{-27}$	$\frac{1}{137}$

Along with gauge bosons acting as force carriers for interactions, most gauge bosons have a degree of self-coupling which leads to self-interactions. The gluon couples with particles that contain-colour charge, but as the gluon itself possesses a colour charge, gluons may interact with other gluons in exchange processes similar to the exchange of a force carrier between two interacting particles. This self-interacting behaviour is also seen in the W and Z bosons as they couple to the weak charge they carry, but no self-interactions are observed for the photon as it does not carry electromagnetic charge [**?**].

1.1.2.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the theory of the strong interaction, mediated by the gluon which couples to colour charge. It corresponds to the $SU(3)_c$ symmetry group of the overall SM. The strong interaction conserves energy, momentum, angular momentum and colour charge. Only quarks and gluons themselves possess colour charge, so quarks are the only fermions to feel the strong interaction. As highlighted above, this also means gluons are capable of self interaction, which leads to two distinct properties of the string interaction: *colour confinement* and *asymptotic freedom*. Colour confinement is the requirement that observable states have net zero colour charge. This means gluons, like quarks, are only observed in bound states. Asymptotic freedom describes how the interaction gets weaker at short distances, and means at close difference such as quark-quark scattering the interaction normally proceeds through a lowest-order single gluon exchange interaction. The converse of this is that the force increases significantly as the interaction distance increases and higher-order Feynmann interaction diagrams become significant [**?**].

1.1.2.2 Electroweak Unification

Electroweak Unification (EW) is the expression of the electromagnetic interaction and the weak interaction as separate manifestations of a combined electroweak force in the Glashow-Weinberg-Salam model [7–9], which corresponds to the $SU(2)_L \times U(1)_Y$ symmetry group. Quantum Electrodynamics (QED) describes the macroscopically observable $U(1)$ electromagnetic force with the photon as the mediating boson, and any interaction conserves energy, momentum, parity and charge and additionally never changes particle type through the interaction. The $SU(2)$ weak interaction is mediated by the charged current vector bosons W^+, W^- and the neutral current vector boson Z, which have large masses that limit the weak interaction to very short distances. The charged current interaction is capable of changing the flavour of a particle and also of violating parity in an interaction.

The weak interaction by itself was observed to diverge from observation at high energies, leading to the introduction of the unified theory. The combined $SU(2)_L \times U(1)_Y$ group produces four gauge bosons which mix to produce the more recognisable γ, W^+, W^- and Z bosons. This weak interaction couples to weak isospin charge, which is an analogous quantity to the colour charge of QCD. As the weak bosons carry weak isospin charge themselves, self coupling of the weak bosons is permitted, but is forbidden for the photon as it does not carry electric charge. The weak interaction has been experimentally observed to violate parity conservation [**?**, **?**].

While the weak interaction acts on both quarks and leptons, weak interaction in the quark sector is affected by *quark mixing*. In this construction, the quark mass eigenstates q participate in weak interactions via the weak eigenstates q' formed from linear combinations of the q states [**?**]. The observable result of this quark mixing is that different flavour changing interactions have different strengths. The coupling relationships of the weak and mass eigenstates is described by the unitary Cabbibo-Kobayashi-Makasawa matrix V_{CKM} [10, 11]

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.1}$$

where shown V_{CKM} here transforming between the two sets of eigenstates q and q'. The CKM matrix elements V_α of V_{CKM} describe the relative couplings of the eigenstates, and are parametrised in terms of three mixing angles and one complex phase [**?**, 11].

1.1.3 Spontaneous Symmetry Breaking: The Higgs Boson

The gauge field theories used for the QCD and EW models when unaltered require massless gauge bosons in order to preserve gauge invariance. This is satisfactory for the gluon and photon, but a separate theory is required to explain the mass of the $W^\pm$ and Z bosons. The *Higgs mechanism* proposed a method for particles to acquire mass by coupling to the spin-0 *Higgs* field via the Higgs boson [12–14]. This process as proposed is an example of a *spontaneous symmetry breaking* process, where the gauge invariance of the interaction is preserved but the ground state breaks the invariance.

The Higgs Mechanism proposed introducing a complex doublet of scalar fields ϕ that interact with the $W^\pm$ and Z fields. In the Lagrangian formulation this results in a term akin to a mass term ($\propto \psi^2$) which effectively links that mass of the bosons to their coupling with this scalar field. This field self interacts to produce a potential energy $V(\phi)$ given by

$$V(\phi) = \mu^2\phi^2 + \lambda\phi^4 \tag{1.2}$$

resulting in an equilibrium point ($\phi = 0$) that respects the symmetry, but is inherently unstable, with an infinite set of degenerate non-zero minima. This minima, the lowest energy level vacuum state occurs at $|\phi^2| = v^2 = \frac{-\mu^2}{2\lambda}$ where the symmetry is *spontaneously* broken. This field, in an analogous fashion to the other quantum fields of the SM, can produce particles from excitations which form the physical *Higgs Scalar Boson H*.

1.2 C

onfirmation of the Higgs boson as part of the SM was only achieved relatively recently [15, 16], where a spin-0 boson consistent with the SM Higgs was observed by the ATLAS and CMS experiments at the LHC. Section **??** covers in more detail the production and behaviour of the Higgs boson in collider experiments.

1.3 Physics of *pp* Collisions

Recent experimental efforts to probe the Standard Model have focused on high-energy collider experiments, where beams of particle with equal energy are collided head on within detector volumns. For proton-proton (*pp*) collisions, matters are complicated as the colliding protons are composite particles, which at high energy consist of the three *valence* quarks and a sea of virtual quarks and gluons. Collectively these constituents are referred to as *partons* where each parton carries a fraction of the overall hadron momentum, and the interaction in the *pp* collision consists of elastic scattering between these partons. At a given energy scale Q^2 the probability that a parton i carries a fraction x_i of the overall momentum is described by the parton distribution function (PDF) $f_i(x, Q^2)$. These PDFs cannot be calculated from QCD but can be determined from experimental measurements, and collections of PDFs have been assembled from the leading collider experiments [17].

In any particle interaction, the probability a particular reaction occurs is in proportion to the cross section of the reaction. The cross section for a short range, hard parton-parton collision is given by $\hat{\sigma}(Q^2)$, where scattering energy scale $Q^2 = x_1 x_2 E_{cm}^2$ in the parton-parton centre-of-mass frame where E_{cm} is the energy in the centre-of-mass frame. To compute the cross section σ for some hard process $pp \longrightarrow X$, all possible combinations of incoming partons must be summed over and the momentum fractions integrated over while accounting for the PDFs,

$$\sigma_{pp\to X} = \sum_{i,j=q,g} \int dx_1 dx_2 f_i(x_1, Q^2) f_j(x_2, Q^2) \hat{\sigma}_{ij\to X}(Q^2) \tag{1.3}$$

1.3.1 Geometry

The high energy protons used in collisions are relativistic in nature, and as the momenta of the colliding partons are not guaranteed to be equal and opposing there is always an unknown element of longitudinal boosting in *pp* collisions. As a consequence, use of light-cone coordinates and some definitions of convenient quanties can be of benefit to *pp* collision analyses [18].

Typically the momentum in the transverse plane p_{T} is used for a particle, and the rapidity y of a particle with non-zero p_{T} is defined:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{1.4}$$

This rapidity y transforms additively to boosts along the z axis, so any rapidity difference between two objects is invariant to such boosts. For cases where the mass of a particle is negligible (highly relativistic particles) the rapidity can be related to the polar angle of the particle as the pseudo-rapidity η:

$$\eta = -\ln \tan \frac{\theta}{2} \tag{1.5}$$

The distance between two objects within the detector is commonly expressed in the (η, ϕ) space rather than absolute, with this separation being given by $\Delta R = \sqrt{(}(\Delta\eta)^2 + (\Delta\phi)^2)$

1.3.2 Collision simulation

A *pp* collision is a complex event which results if a significant number ($O(1000)$) of final state particles, each of which interact and evolve over the timescale of an event. This progression of the collision event can be broken down into distinct stages of behaviour of the produced particles: the *hard process*, *parton shower*, *hadronisation*, *unstable particle decays* and *underlying event.*

This breakdown is key to the simulation of *pp* collisions using Monte-Carlo event generators, the use of which is critical in current high energy physics research. Monte-Carlo simulations of collisions are used to predict and prepare for real data-taking experiments, obtain control datasets of particular particle interactions and act as controls to optimise analysis tools. The breakdown of the interaction into distinct stages has allowed specialised software to be produced for each step, which makes use of a characteristic scale and certain safe approximations for the step to provide reliable predictions, while reducing the computational demands of the simulation [19].

1.3.2.1 Hard Process

The first stage of a *pp* collision and the first step of a simulation, the hard scatter refers to highest momentum transfer process in the event between coloured particles, and forms the core of the event. This details the interaction of partons entering the event and those outgoing partons resulting from the process. In simulation the probability distribution of the partons is calculated from perturbation theory to the desired accuracy (LO, NLO etc) using the PDFs of the constituents.

1.3.2.2 Parton Shower

While the hard scatter interaction in a collision is relatively straightforward, the overall behaviour of the partons is much more complex as they progress through the event. The incoming and outgoing partons from the hard scatter radiate additional interaction particles during the event The Bremsstrahlung radiation of photons by scattered electric charges is well described by QED, and the analogous radiation of gluons by scattered colour charges as explained by QCD produces additional partons within the interaction. However, as the gluons produced by QCD scattering themselves carry colour charge, there is extending showering of gluons producing gluons, resulting in the phase space of the interaction being filled with a sea of soft gluons. Both of these radiative processes make up the parton shower stage of the event simulation.

The evolution of these parton showers is evaluated in Monte-Carlo simulations using a step-by-step iterative process, on the scale of momentum transfer in the interaction. This process is started at the hard scatter and evolved through the interaction with decreasing momentum scale until the point at which perturbation theory breaks down, necessitating a different evaluation method.

1.3.2.3 Hadronisation

With the breakdown of perturbation theory at low momentum scales, observable colourless hadrons are constructed from the coloured partons using hadronisation models in order to extend the simulation. These hadrons are the physical final state particles observed in the detector, which exist due to the colour confinement of the quarks and gluons. Within a particle detector, rather than individual hadrons *jets* of hadrons are observed. As a coloured fragment produced in the interaction moves away from the interaction, it will create other coloured fragments around itself in order to produce a confined hadron while moving away from the collision. This occurs repeatedly for each hadron ejected from the collision as it moves away, producing a collimated stream of hadrons which make up a jet.

In simulation this step involves collecting the partons produced in the parton shower into hadrons, and is typically evaluated using either a String model [20] or a Cluster model [21]. These steps are models, and not calculations as to how the partons combine, as to calculations being prohibited by the breakdown of perturbation theory.

1.3.2.4 Unstable Particle Decays

The final stage of the evolution of the parton shower considers the hadrons produced during the shower. These hadrons may not be stable particles but could be resonances that go on to decay within the detector to produce the more stable hadrons observed in the data. Most modern simulation software models these decays, but the exact specification of the decay tables and channels has a significant impact on the final state of the simulation.

1.3.2.5 Underlying Event

While the hard scatter and subsequent parton shower results from the highest momentum interaction of the *pp* collision, the remnants of the proton not involved in this will continue to interact with each other. This produces additional soft hadrons that fill the interaction environment, overlapping with the products of the hard scatter interaction.

The dominant model for simulation of the underlying event is a perturbative model where the other components undergo additional discrete hard scatter interactions and corresponding parton showers which are simulated in an corresponding fashion to the core scattering.

1.3.3 Monte-Carlo Software

There is a broad selection of software tools for evaluating *pp* collisions, from general purpose simulations like Pythia [22] or Sherpa [23] which are used to evaluate the complete process, to more specific tools like Powheg [24] which is used to produce hard scatter events with NLO matrix elements. Most software packages make use of the chain of generation for an event outlined previously, and modern analyses will make use of multiple generators interfaced together to compute different steps with improved accuracy.

1.4 The Higgs Boson

Detecting the SM Higgs boson is strongly dependent on the predominant production and decay channels for the Higgs boson, which in turn depend on the specifications of the collider used for the search. In this section the relevant production and decay channels at the Large Hadron Collider (LHC) will be discussed.

1.4.1 Higgs Production

While there are many various methods for production of a Higgs boson, at the LHC the cross section is dominated by gluon-gluon fusion ($gg \rightarrow H$) as shown in Figure **??**, with the second largest cross-section arising from Vector Boson Fusion (VBF, Section **??**). Other significant production processes are the associated production with a weak boson (WH/ZH, Higgs-strahlung) production modes and associated production with top quarks (ttH) [1]. The lowest order Feynmann diagrams for these processes are shown in Figure **??**.

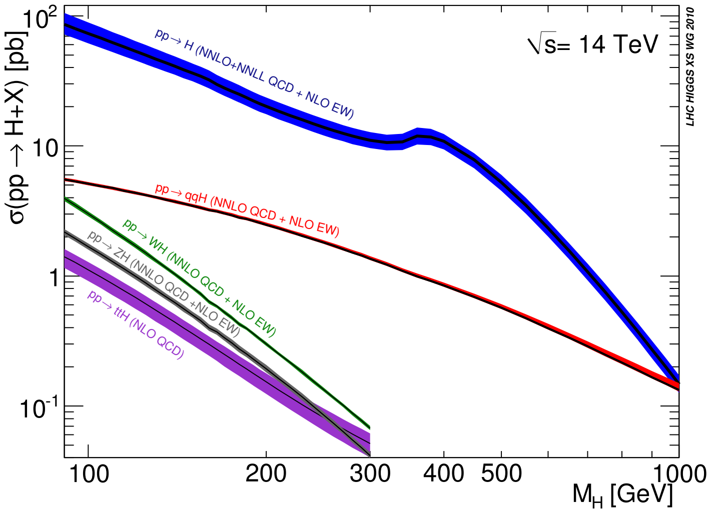

Figure 1.1: SM Higgs Production cross section for $\sqrt{s}$ = 14 TeV. $pp \longrightarrow H$ corresponds to gluon-gluon fusion production and $pp \longrightarrow qqH$ vector boson fusion. [1]

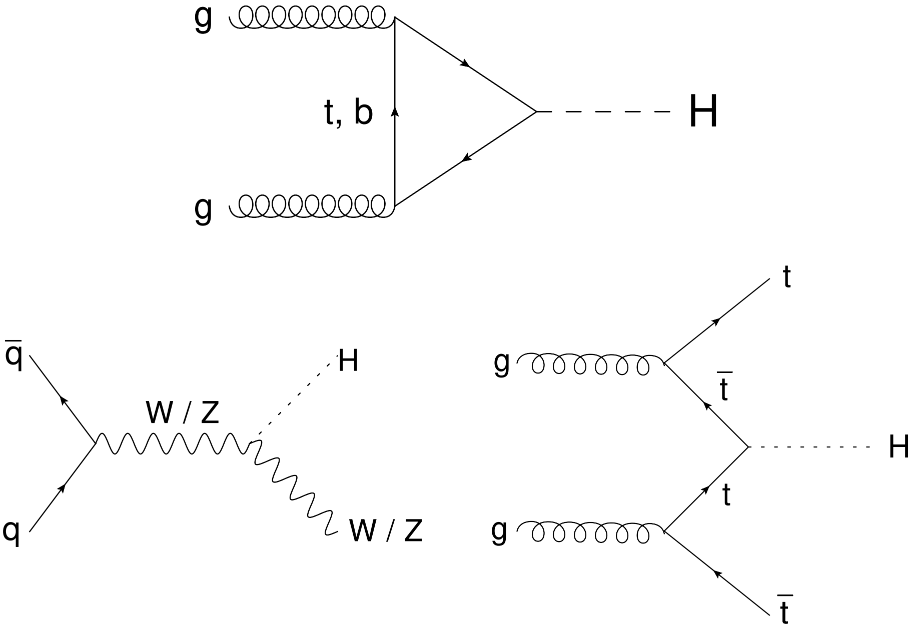

Figure 1.2: Lowest order Feynmann diagrams for gluon-gluon fusion ($gg \longrightarrow H$), W/Z associated production (WH/ZH) and top anti-top associated production ($t\bar{t}H$).

1.4.1.1 Gluon-gluon Fusion

The dominant production mechanism for the Higgs boson in hadron colliders is the $gg \longrightarrow H$ production via in intermediate quark loop. The dynamics of this mechanism are controlled by strong interactions, thus calculations of QCD corrections are necessary for any accurate predictions, and have been computed up from next-to-leading order (NLO) to N^{3}LO for the

$gg \rightarrow H$process in recent years, along with the inclusion of Electro-Weak corrections in the cross section calculations [1].

1.4.2 Vector Boson Fusion

Production of a Higgs boson from the fusion of vector bosons radiated from initial-state quarks is the second largest cross-section at the LHC, and is useful as a production mode due to topological characteristics which can distinguish the event from $gg \rightarrow H$. In VBF $H \rightarrow b\bar{b}$, the characteristic topology is a pair of central b-jetsforming the Higgs candidate, and two forward, close to the beam line VBF jets formed from remnants of the initially colliding protons as displayed in Figure **??**. In addition central jet activity is suppressed due to the lack of colour exchange between the colour single Higgs boson and the decay b-quarks [25]. These distinct features mean that while the cross section for VBF at a Higgs mass of < 200 GeV is dominated by $gg \rightarrow H$, the easy to detect signature means the channel is a cornerstone of searches for the Higgs boson.

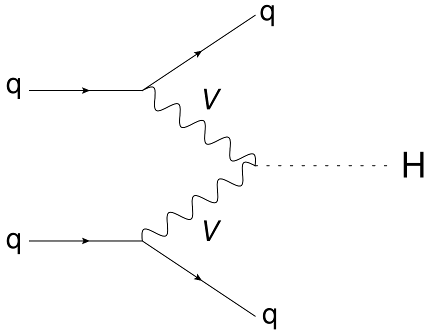

Figure 1.3: Feynmann diagram for the production of a Higgs boson via Vector Boson Fusion, where q denotes any quark or antiquark

1.4.3 Higgs Decay

The branching ratios for decays of the Higgs boson in the Standard Model have been extensively determined using Monte-Carlo event generators. As is to be expected, the relative cross-sections of the decay modes are strongly dependent on the mass of the Higgs boson, as highlighted in Figure **??**.

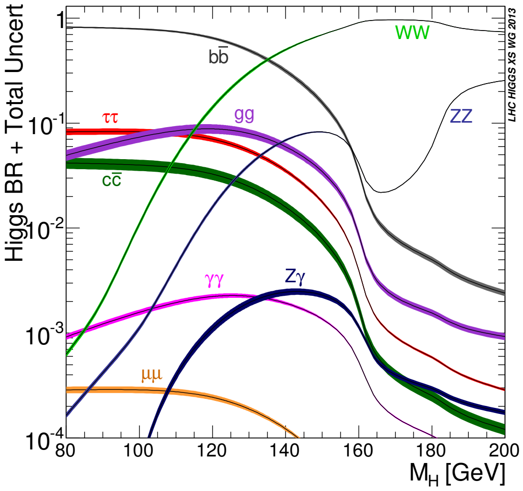

Figure 1.4: Higgs decay branching ratios for the low mass region with their uncertainties [2].

While observations consistent with the Standard Model Higgs boson have been made for the $H \longrightarrow \gamma\gamma$, $H \longrightarrow ZZ$, $H \longrightarrow W^+W^-$ and $H \longrightarrow \tau^+\tau^-$ channels, observation of th $H \longrightarrow bb$ decay channel is significantly hindered owing to the large background from multijet production in hadron collisions. Despite this, the topology of the VBF production mechanism makes it a viable option for observation of the $b\bar{b}$ decay channel.

1.4.4 VBF Searches

Searches fo the VBF $H \longrightarrow b\bar{b}$ interaction look for a resonance in the invariant mass of a pair of jets containing b-quarks(m_{bb}) in events with the characteristic topology. This characteristic topology distinguishes the signal events from the multijet events that form the dominant background with a non-resonant m_{bb} spectrum. An additional resonant background contribution to the m_{bb} spectrum is due to decay of a Z boson to two jets in association with two jets.

In the most recent searches for the Higgs boson produced via VBF, which this analysis emulates, the VBF $H \longrightarrow b\bar{b}$ events are indistinguishable from the $gg \longrightarrow H$ events, and are separated using a multivariate boosted decision tree (BDT) analysis to refine the phase space to the most VBF sensitive BDT regions.

Chapter 2

Detector

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular particle accelerator operated at European Organisation for Nuclear Research (CERN, Conseil Européan pour la Recherche Nucléaire). Currently the largest accelerator in the world, the LHC is designed to collide opposing beams of protons at a *centre-of-mass* energy $\sqrt{s} = 14\text{TeV}$ and a peak *luminosity* of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ [26]. The first proton beams were circulated in the LHC in 2008, with Run-1 of LHC data taking being conducted from 2010 to 2013 at increasing $\sqrt{s}$ of 0.9, 7, and 8TeV, after which the machine was shut down for scheduled maintenance. Following on from the long shut down period, Run-2 of the LHC has been ongoing since 2015, operating at $\sqrt{s} = 13\text{TeV}$.

The principal LHC ring consists of eight pairs of alternating long arc sections and short straight insertion sections, situated within the underground tunnel excavated for the older Large Electron Positron Collider experiment [27, 28]. The arc sections contain the dipole magnets used to bend the particle beam around the ring, while the straight sections contain four interaction points, at each of which the large experiments are located. The remaining straight sections contain the operational systems of the LHC: beam acceleration, injection, dumping and collimation. The proton beams are generated outside the principal ring and inserted into the ring by the LHC injector chain, a sequence of smaller accelerators which are used to bring the proton beams up to a suitable energy for injection. The proton beams injected into the accelerator are obtained from a cloud of hydrogen gas, which is passed through an electric field to strip the electrons before the protons are inserted into the beam acceleration components. The proton beams are arranged such that the protons move in bunches of $O(10^{11})$ protons, with multiple bunches placed into trains. During Run-2 the LHC operated with bunch spacings of 50ns and 25ns between the bunch trains.

The principle measure of the operation of the LHC is the instantaneous beam luminosity L. This parameter is a measure of the rate of collisions within the accelerator, given by

$$L = \frac{1}{\sigma}\frac{dN}{dt} = \frac{n_b n_1 n_2 f}{2\pi \Sigma_x \Sigma_y} \tag{2.1}$$

where in the general case σ is the interaction cross section, $\frac{dN}{dt}$ is the event rate, n_b, n_1 and n_2 are the number of bunch crossing producing collisions, and the number of bunches in both of the colliding beams, f the machine revolution frequency, and $\Sigma_{x,y}$ are parameters relating to the beam width. This instantaneous luminosity is integrated across a time period, such as an LHC Run or a specific data period, to produce the integrated luminosity $\int Ldt$ which is a measure of the total recorded data.

Once a beam is accelerated to the target energy, collisions begin at the interaction points. Interactions are ongoing for periods of several hours, and will go on until the the beam is replaced due to general decay of the interaction rate or beam instabilities.

At the LHC, the four large experiments at the interaction points are ATLAS (A Toroidal LHC ApparatuS), CMS (Compact Muon Solenoid), LHCb (LHC beauty) and ALICE (A Large Ion Collider Experiment). LHCb is a forward spectrometer heavy flavour experiment, designed to study flavour physics with emphasis on the b-quark and on matter/anti-matter asymmetry. ALICE focuses on the collisions of heavy ions, while ATLAS and CMS are general purpose detectors to conduct experiments across a broad range of modern physics research areas.

2.1.1 LHC Run Conditions in 2016

Over the course of 2016, following beam commissioning runs, the LHC beam was operated predominantly with two beams of energy 6.5TeV for $\sqrt{s} = 13$TeV. Over the course of the 2016 data-taking the LHC provided an integrated luminosity of ~ 40 fb^{-1} to the ATLAS and CMS experiments with a peak instantaneous luminosity of 1.4×10^{34}cm^{-2}s^{-1} with 2220 bunches per beam [29].

2.2 The ATLAS Detector

The ATLAS detector [6] is a multi-purpose detector designed to study a broad selection of physics phenomena within the experimental conditions of the LHC. The detector is cylindrical in structure with the axis aligned to the beam path and nominally forward-backward symmetric in terms of the beam collision point at the centre of the detector. The detector provides approximately 4π solid angle coverage around the interaction point to detect as many collision products as possible.

The structure of the ATLAS detector is composed of concentric subsystems around the interaction point. The Inner Detector (ID) is the component closest to the interaction point, and is contained in a superconducting solenoid. This is surrounded by high-granularity calorimeters and an extensive muon spectrometer contained within an eight-fold azimuthally symmetric arrangement of three large toroidal magnets. A schematic representation of the ATLAS detector is shown in Figure **??**. The detector consists of three main sections, two *endcaps* located on the ends of the detector and a central *barrel* section. A summary of the operational parameters of the principle detector components is given in Table **??**.

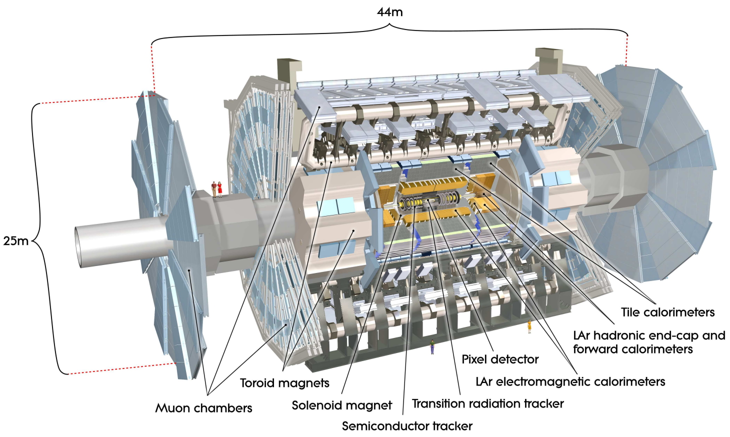
44m
25m
Tile calorimeters
LAr hadronic end-cap and forward calorimeters
Pixel detector
LAr electromagnetic calorimeters
Toroid magnets
Solenoid magnet
Transition radiation tracker
Muon chambers
Semiconductor tracker

Figure 2.1: Schematic cut-away of the ATLAS detector [3].

The conventional coordinate system used to describe the detector takes the interaction point as the origin, with x pointing horizontally out into the centre of the detector ring, y out and upwards with z along the direction of the beam line. The angle ϕ describes azimuthal rotation around the beam pipe and θ is the polar angle along the beam line.

2.2.1 Inner Detector

The Inner Detector [30] (ID) provides pattern recognition, momentum measurements, electron identification and measurements of both primary and secondary vertices to efficiently identify b-hadron decays within a pseudorapidity range $|\eta| < 2.5$. The ID itself is contained within a 2 T solenoidal field, which is used to bend the paths of charged particles within the ID. The ID is specifically designed to have a high momentum resolution (Table **??**), and consists of three separate detector sections: the silicon pixel detector provides fine granularity track and vertex reconstruction, the silicon strip semiconductor tracker measures the trajectory of transiting charged particles and the outer transition radiation tracker used for particle identification is comprised of layers of straw tubes containing mixtures of xenon, oxygen and carbon dioxide [6].

2.2.2 Calorimeters

Calorimeters are used to measure the energy of interacting particles moving out from the interaction point. These particles cause the development of energy showers within the calorimeter substrate, forming different shower types depending on the interaction force of the particle, with electromagnetic (EM) showers forming from EM interactions and hadronic showers forming from interactions via the strong nuclear force. The energy deposited in this shower can be then used to calculate the energy of the incoming particle. The ATLAS calorimetry system consists of a combination of EM and hadronic calorimeters arranged with full ϕ-symmetry around the beam axis. The combination of all separate calorimeters provides pseudorapidity coverage in the range $|\eta| < 4.9$. Within the pseudorapidity region of the inner detector, the fine granularity of EM calorimeters is optimised for measurements of electron and photons, while the coarser hadronic calorimeters contained in the remainder of the calorimeter system are sufficient for measurements of the energy of produced hadrons. The structure and design of the calorimeter components has been optimised to provide complete azimuthal coverage, take into account the engineering requirements for assembling the detector and account for radiation considerations between the different detector components [6].

The EM calorimeter [31] is a lead-Liquid-Argon (LAr) detector, which is split into a barrel section (EMB, $|\eta| < 1.475$) and two endcap sections (EMEC, $1.375 < |\eta| < 3.2$) with each section contained in a separate cryostat. The EMB consists of two identical half-barrels split by a small gap at $z = 0$. Each of the EMEC sections is a pair of coaxial wheels, with the inner and outer sections covering regions $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$ respectively. The major body of the EM calorimeter is divided into three sections of decreasing cell granularity, moving out from the beamline.

Hadronic calorimetry for particles undergoing the strong interaction is provided by the steel/scintillator tile calorimeter [32] for pseudorapidity values of $|\eta| < 1.7$, and by the LAr flat-plate Hadronic Endcap Calorimeter (HEC) for $1.5 < |\eta| < 3.2$. The tile calorimeter directly surrounds the EM calorimeter, and is split into a central barrel section for $|\eta| < 1.0$ and two extended barrel sections covering $0.8 < |\eta| < 1.7$. The HEC, akin to the EMEC, consists of two separate wheels per end-cap covering $1.5 < |\eta| < 3.2$, and is contained within the same cryostat as the EMEC. The HEC consists of alternating copper plates with LAr gaps to act as the active medium.

In addition to the barrel and end-cap calorimeters, the LAr Forward Calorimeter [33] is contained within the end-cap cryostat (The FCal is omitted from Figure **??**) and is designed to perform both EM and hadronic calorimetry across a pseudorapidity range of $3.1 < |\eta| < 4.9$ using a combination of copper/LAr (EM) and tungsten/LAr (hadronic) calorimeter components.

2.2.3 Muon Spectrometer

The muon spectrometer is the outermost component of the ATLAS detector, measuring trajectory and momentum of muons from the interactions within a pseudorapidity range of $|\eta| < 2.7$. The muon system consists of three large superconducting coils that deflect the muon trajectories and a suite of tracking devices. The system is designed for high precision tracking of the minimally ionising muons and for use in the triggering system of the overall detector. The triggering chambers consist of Resistive Plate Chambers which can respond to a particle transit in $\mathcal{O}(10)$ns, while the precision momentum measurement is carried out in Monitored Drift Tubes arranged in layers [6].

Table 2.1: Performance goals and operational ranges for the principal components of the ATLAS detector. [6]

System	Component	η Coverage	Resolution
Tracking		$0 < \|\eta\| < 2.5$	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.05\% p_\mathrm{T} \oplus 1\%$
EM Calorimetry	EMB	$0 < \|\eta\| < 1.475$	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$
	EMEC (Inner)	$1.375 < \|\eta\| < 2.5$	
	EMEC (Outer)	$2.5 < \|\eta\| < 3.2$	
Hadronic Calorimetry	Tile (Barrel)	$0 < \|\eta\| < 1$	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$
	Tile (Extended)	$0.8 < \|\eta\| < 1.7$	
	HEC	$1.5 < \|\eta\| < 3.2$	
Forward Calorimetry	FCal	$3.1 < \|\eta\| < 4.9$	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$
Muon Spectrometer		$0 < \|\eta\| < 2.7$	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 10\% \, at \, p_\mathrm{T} = 1$ TeV

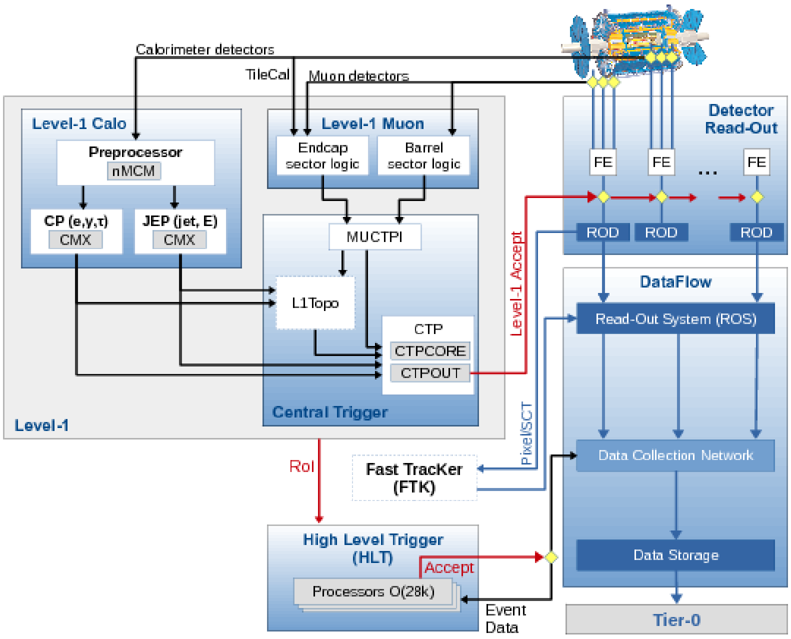

Figure 2.2: Schematic plot of the ATLAS Trigger and Data acquisition system [4].

2.3 Trigger and data acquisition

When operating at the design luminosity, the LHC produces a bunch-crossing rate of 40 MHz [34]. This extreme rate of interaction necessitates a trigger system to reduce the output rate to a suitable level for offline processing, which is predominantly limited by the rate at which data can be written to disk. The trigger system selects events by quickly identifying distinguishing features of events, signatures of muons, electrons, jet and b-jet objects, and using combinations of these signatures to signify an event as relevant for further analysis.

The ATLAS trigger system consists of a chain of selection stages of increasing severity and corresponding decrease in rate. A schematic outline covering both the logical process and the transfer of data between components of the trigger chain is shown in Figure **??**. The principal decision logic of the trigger system is contained in two sections, the Level 1 (L1) trigger system and the High Level Trigger (HLT).

The L1 trigger system [35] is a hardware-based decision system, using fast custom electronics to minimise latency in any decision. The L1 uses reduced-granularity data from the calorimetric and muon detectors, reconstructed objects and missing and total transverse energy. The high bunch-crossing rate means instantaneous processing of the event is non-viable, so event readouts are stored in a buffer chain of events to be evaluated with a fixed permitted decision time per event. Along with this first selection, the L1 trigger defines

Regions of Interest (RoIs) in the phase space within the detector, which are labeled for investigation in the HLT.

In contrast to the hardware computation of the L1 system, the HLT consists of software algorithms running in a farm of $\approx$ 40000 interconnected processors [34]. Following acceptance of an event by the L1 trigger, events are transferred from the initial data pipeline to dedicated readout buffers for the HLT. The HLT performs processing on the events using finer-granularity information from the calorimeters and muon spectrometer, along with making use of information from the ID, which is unavailable to L1. This more precise data is then computed using object reconstruction algorithms to generate particle objects similar to the objects reconstructed at a later point the data has been stored. The decision at HLT level to store an event is managed by a trigger chain, which is a sequence of specific criteria and algorithms evaluated on an event in sequence.

A key component of the trigger chain is the prescaling factor of the chain, where the overall output rate of the trigger chain is reduced by the prescale factor to bring the output rate within bandwidth limits. The trigger menu in 2016 provided a selection of main ATLAS triggers used for the data-taking [36], with $O(1000)$ independent HLT trigger chains for evaluating events. Along with the partial reconstruction of relevant objects, the HLT is capable of performing complete reconstruction of an event, and also capable of writing out these partial or complete reconstructions of an event into different data streams from the complete detector readout for use in analysis. The standard terminology for events and data recorded and processed during the operation of the LHC is *online* data, while objects and information produced by considering the output of the detector after the data has been stored is termed *offline*. These terms are used extensively throughout the rest of this thesis to distinguish between the different data sources.

Overall usage of the trigger system brings the output rate down to 1 kHz with a maximum L1 trigger rate of 100 kHz.

2.4 Event Cleaning

Beyond the reduction in the event storage rate handled by the trigger chains and prescaling, only select sections of the overall data output by the LHC are ever used in analyses. The LHC is not free from operational errors or issues with the hardware and software of the detector. Parts of the output data can be corrupted by incomplete events due to detector failings, poor data integrity or disruption of the machine. From the complete output for a Run section, which is divided into luminosity blocks, only the blocks which have been marked as *good* are made use of in analyses. The internal directory of usable luminosity blocks is named the Good Runs List (GRL).

Along with these event selections based on using correct data, analyses typically refine events down to a particular area of focus, which is discussed in Chapters **??**, **??** and **??**.

2.5 Object Reconstruction

2.5.1 Jets

As discussed in Section **??** the high p_{T} quarks and gluons produced during *pp*-collisions result in collimated streams of hadrons called jets, which are the physical objects detected in the event. Detectors make use of algorithms to reconstruct these jets from the calorimeter readouts to relate the stream of hadrons to the initial fragmented partons. There are various algorithms used to reconstruct jets within the ATLAS detector, and these algorithms commonly require the definition of a jet to be invariant under additional soft or collinear emissions. Such algorithms are designated as infra-red (IR) or collinear (C) safe.

Modern jet algorithms are broadly split into two types: cone-type and sequential clustering algorithms. Cone-type algorithms take the hardest (highest momentum) object in an event as a seed of an iterative process of looking for a stable cone rooted at this seed [37]. Once a cone is defined, any constituents contained within the cone are removed from consideration and the process repeats. The alternative sequential clustering algorithms assume that particles within jets will have small differences in transverse momentum and groups particles based on the momentum space to reconstruct the jets. Sequential clustering algorithms function using iterative steps with two distance parameters. The first distance is the separation between two particles d_{ij}, defined as

$$d_{ij} = min(p_{\mathrm{T}i}^{a}, p_{\mathrm{T}j}^{a})\frac{\mathit{\Delta} R_{ij}^{2}}{R} \tag{2.2}$$

where a is a particular exponent for a given algorithm, R is the radius parameter of the final reconstructed jet size and $\mathit{\Delta} R_{ij}$ is the (η, ϕ) space distance between the two objects. The second parameter d_{iB}, is the momentum space distance between the beam axis and an object [38] and is given by

$$d_{iB} = p_{\mathrm{T}i}^{a} \tag{2.3}$$

The principal algorithm used for jet reconstruction at ATLAS is the anti-k_t algorithm [39], which is a sequential clustering algorithm with $a = -2$. The algorithm is seeded with the highest p_{T} particles in the event, and iteratively computes the distance parameters. At each step, the two are compared: if d_{ij} is smaller, particles i and j are combined whereas if d_{iB} is smaller particle j is labeled as a jet. The fact this algorithm tends to result in approximately circular reconstructed jet objects makes it favourable for experimental analyses as they are easily calibrated. The anti-k_t algorithm is IRC safe and typical used with $R = 0.4$ in the ATLAS experiment, and can be readily applied to clustering partons and calorimeter deposits in addition to hadrons.

During jet reconstruction, when the energy deposits are extracted from the calorimeter, there is the option of reading the calibrated [40] calorimeter cells according to the Electromagnetic (EM) scale, or by applying Local Cell (LC) corrections [41] to account for the attenuated physical response of the calorimeter and the difference in hadronic and electromagnetic response, which restores the energy of extracted objects to correspond to Monte-Carlo simulated truth objects. In this analysis, readouts of all jet objects, both offline and trigger level, were taken at the EM energy scale.

2.5.1.1 Pileup

As mentioned in Section **??** on the process of a *pp* collision, there are significant interactions as a result of the parton interactions accompanying the hard-scatter interaction of the collision. In addition to this underlying event, additional *pp*-collisions within a particular bunch crossing will contaminate the event. The collection of these jets from other *pp*-collisions in the detector output is termed in-time pileup [42]. In addition to the in-time pileup, interactions from preceding or subsequent bunch crossings also contribute contaminating objects to the detector readout, which is named out-of-time pileup. In-time and out-of-time pileup are collectively referred to as pileup in the detector, and necessitate processing and calibration of the detector output to remove the effects from consideration [43].

2.5.2 *b*-Tagging

Hadrons containing a *b*-quark tend to feature a signature topology as a result of the long lifetime of *b*-hadrons. The extended lifetime results in a significant mean flight path of the *b*-hadron between its production and decay, forming a displaced secondary vertex from the primary hard scatter interaction point. This distinctive structure can be used to identify *b*-jets, and algorithms that exploit this are known as lifetime-based tagging algorithms [44].

Identification of jets containing in ATLAS is based on combining the output of three separate lifetime-based *b*-tagging algorithms [45]: Impact Parameter based algorithms (IP2D and IP3D, Section **??**), Secondary Vertex based (SV, Section **??**) and Decay Chain based (JetFitter, Section **??**) into a multivariate discriminant (MV2, Section **??**) which is used to distinguish the jet flavours. These algorithms have undergone continuous improvement over the Run-2 cycle of the LHC to improve the separation of jet flavours.

The inputs for each of the b-tagging algorithms are all taken from the ID of the ATLAS detector (Section **??**). This limits b-tagging to jets with $|\eta| < 2.5$, and in addition jets with a $p_{\mathrm{T}} < 20$GeV are not selected for b-tagging, nor jets determined to be likely a result of pileup in the detector which are eliminated using a multivariate discriminant from Jet Vertex Tagger algorithm [43, 46].

2.5.2.1 IP2D and IP3D: Impact Parameter based Algorithms

To identify b-hadron decays, impact parameters of tracks from the secondary vertex can be computed with respect to the primary vertex of the interaction. The IP2D algorithm uses a transverse impact parameter d_0 defined as the distance of closest approach of a track to the primary vertex in $(r,\ \phi)$ plane around the vertex. The IP3D algorithm uses both the transverse impact parameter and a correlated longitudinal impact parameter $z_0 \sin\theta$, defined as the distance between the point of closest approach in $(r,\ \phi)$ and the primary vertex in the longitudinal plane [47]. These parameters typically have large values as a result of the lifetime of b-quark. The signs of the impact parameters are also defined to take account of whether they lie in front or behind the primary vertex with respect to the jet direction, with secondary vertices occurring behind the primary vertex normally due to background.

The significance of the impact parameter values $(\frac{d_0}{\sigma_{d_0}}, \frac{z_0}{\sigma_{z_0 \sin\theta}})$ for each track are compared to probability density functions obtained from reference histograms derived from Monte Carlo simulation, with each track being compared to a selection of reference track categories. This results in weights which are combined using a log-likelihood ratio (LLR) discriminant to compute an overall jet weight separating the b, c, and light-jet flavours from each other. [44, 46]

2.5.2.2 SV1: Secondary Vertex Finding algorithm

The secondary vertex algorithm uses the decay products of the b-hadron to reconstruct a distinct secondary vertex [47]. The algorithm uses all tracks that are significantly displaced from the primary vertex associated with the jet, forming vertex candidates for all pairs of track, while rejecting any vertices that would be associated with decay of long lived particles (e.g. Λ), photon conversions or interactions with the material in the detector. The tracks forming these vertex candidates are then iteratively combined and refined to remove outliers beyond a χ^2 threshold leaving a single inclusive vertex.

The properties of this secondary vertex are used to differentiate the flavour of the jet. The SV1 algorithm is based on a LLR formalism similar to the IP algorithms, and makes use of the invariant mass of all charged tracks used to reconstruct the vertex, the number of two track vertices and the ratio of the invariant mass of the charged tracks to the invariant mass of all tracks. In addition the algorithm is signed in a similar fashion to the IP algorithms and uses the ΔR between the jet direction and secondary vertex displacement direction in the LLR calculation. The algorithm uses distributions of these variables to distinguish between the jet flavours [44, 46].

2.5.2.3 JetFitter: Decay Chain Multi based Algorithm

The JetFitter algorithm exploits the topological structure of weak b-hadron and c-hadron decays inside the jet to reconstruct a full b-hadron decay chain. A Kalman filter is used to find a common line between the primary, b-hadron and c-hadron vertices to approximate the b-hadron flight path [48]. A selection of variables relating to the primary vertex and the properties of the tracks associated with the jet are used as input nodes in a neural network. This neural network uses the input variables, p_{T} and $|\eta|$ variables from the jets, reweighted to ensure the spectra of the kinematics are not used in the training of the neural net. The neural network outputs discriminating variables relating to each jet flavour which are used to tag the jets [44].

2.5.3 Multivariate Algorithm

The output variables of the three basic algorithms described prior are combined as input into the Multivariate Algorithm MV2. MV2 is a Boosted Decision Tree (BDT) algorithm (Appendix B) which has been trained on $t\bar{t}$ events to discriminate b-jets from light and c-jets. The algorithm makes use of the jet kinematics in addition to the tagger input variables to prevent the kinematic spectra of the training sample from being used as discriminating factor. The MV2 algorithm is an revised version of the MV1 algorithm used during Run-1 of the LHC, and has three sub-variants (MV2c00, MV2c10, and MV2c20) of the algorithm distinguished by the exact background composition of the training sample. The naming convention initially referred to the c-jet composition of the training sample; for MV2c20 the b-jets are designated as signal jets where a mixture of 80% light jets and 20% c-jets was designated as background [45].

The MV2 algorithm has a set of working points, defined by a single value of the output distribution of the algorithm, which are configured to provide a specific b-jet selection efficiency on the training $t\bar{t}$ sample. Rather than being used independently, physics analyses will make use of several working points as an increase in b-jet efficiency (corresponding to *looser* b-jet selection) will bring an increased mistag rate of light and c-jets.

These algorithms were refined prior to the 2016 Run-2 data-taking session in response to c-jets limiting physics analyses more the light-jets. This change to enhance the c-jet rejection meant that for the MV2c10, the c-jet fraction was set to 7% in training and the fraction for MV2c20 was 15%. There were a selection of other improvements made to the algorithm relating to the BDT training parameters and the use of the basic algorithms before the 2016 data taking. With these refinements, the MV2c10 algorithm was found to provide a comparable level of light-jet rejection to the original 2015 Mv2c20 algorithm with improved c-jet rejection, so was chosen as the standard b-tagging algorithm for 2016 analyses [46].

2.6 Trigger-Object Level Analysis

In physics analyses at the LHC, the 1kHz event readout rate to storage is significantly below the 40MHz bunch crossing rate. This bottleneck is caused by the limited bandwidth (event rate × event size in bytes) available to analysis channels. In searches with large backgrounds or those with low rates, the prescaling introduced in the trigger system critically affects the amount of significant events output to storage, limiting the statistical power of any search in these hard to isolate channels as a large number of events are discarded to keep output within bandwidth limitations.

This constraint can be alleviated by recording only a fraction of the detector readout for any given event, specifically the jet information reconstructed by the triggering system. This partial event corresponds to a reduction in the event size in bytes which allows for present bandwidth limitations to be upheld with an increased event rate. This process of using the objects produced in the trigger as substitutes for the offline objects is referred to as Trigger-Object Level Analysis (TLA) [49].

In these analyses, partially built events are collected using an additional TLA stream of the output data, which records the jet four-momentum along with a selection of additional identifying variables for jet objects in the HLT, triggered by jet objects from the L1 trigger. The readout does not include individual calorimeter cells nor information from the muon or tracking detectors, and in prior application of a TLA approach to a search for light dijet resonances [49] a partial TLA event was 5% of the size of a full detector readout, and TLA events were read out frmo the detector at a rate of 2kHz.

Chapter 3

Event Selection

This chapter is describes the selection criteria for data and simulated Monte-Carlo events, along with the specific calibrations and configurations used in the extraction and reconstruction of the objects making up the analysis. The event selections described here were chosen to target the typical VBF $H \to b\bar{b}$ final state topology described in Section **??**.

3.1 ATLAS Event Data

The raw data from the ATLAS detector is stored in a proprietary data format used by the ATLAS experiment, the Analysis Object Data (AOD) format. This is the output of the event reconstruction software, with each event having a corresponding discrete entry. For Run-2 of the LHC experiment, this was upgraded to the xAOD format, which is readable by ROOT [50], a modular software framework managed by CERN and designed specifically for analysis of large datasets with complex statistical analysis, visualisation of data and storage. The xAOD format is a many leveled branching tree structure, with nodes of the tree grouping together related information from each event, and has an associated Event Data Model (EDM) to standardise classes, interfaces and types for representation of an event facilitating simple analysis [51].

Analyses typical make use of a derivation framework to refine the complete xAOD into a more selective Derived xAOD (DxAOD) which will normally only the relevant objects to a target analysis, and results in a smaller dataset that is much easier to manipulate, store and operate over. These derivations are produced using the ATLAS bulk data processing framework Athena [52]. The computation framework used for analysis of the xAOD data is the internally developed AnalysisBase suite of tools. The analysis presented in this dissertation uses AnalysisBase Release `2.4.31` and made use of the EventLoop package for event processing.

This set of tools is used for both the real event data and the simulated Monte-Carlo data, with DxAODs of both datasets forming the core data for any ATLAS physics analysis. These datasets, following from the large output rate of the LHC, are extremely large, necessitating the use of parallelised computation to perform any statistically significant analysis. The computational framework developed at ATLAS is designed to perform concurrent computation, and processing, making use of the Worldwide LHC Computing Grid [53] to provide the necessary hardware capacity.

3.2 Datasets

The proton-proton collision data was recorded in 2016 at a centre-of-mass energy of $\sqrt{s}$ = 13TeV. In this dissertation, Data Period D was used owing to limited storage space on analysis computing facilities. For events from the ATLAS detector to be considered usable for analysis, there are certain quality criteria that need to be passed by the event. Events are subdivided into luminosity *blocks*, which are marked as *good* if there are no flaws in the data integrity or missing information from the detector readout. The events were marked as *clean* if there were no errors reported for the tracker or calorimiter components of the detector, and only clean events were studied.

Information on whether certain luminosity blocks are marked as clean are contained within a configuration *Good Runs List* (GRL).This analysis used the all year 25ns Good Runs List (Table A.1, Appendix A), resulting in a data luminosity of 4.6312fb^{-1}.

3.3 Monte-Carlo simulated events

The simulated VBF sample (Table A.1, Appendix A) was produced during the MC15c production period. This sample was produced using the NLO generator POWHEG configured using the CTEQ6L1 [54] set of PDFs and interfaced with PYTHIA8 tuned to AZNLO [55]. The response of the ATLAS detector to the Monte-Carlo events was simulated using the GEANT4 [**?**, **?**] simulation, which recreates a configurable model of the ATLAS detector, and performs the same calibrations and reconstructions on the Monte-Carlo simulated events as the physical detector performs on data.

In order to accurately compare the simulated events from the Monte-Carlo samples with the real event dataset, it is necessary to normalise the Monte-Carlo samples to the total luminosity of the dataset, based on the theoretical cross-section for the interaction. The Monte-Carlo simulation assigns a weight w_i to each event simulated, which are summed to give the total number of events in the Monte-Carlo. Each bin of any histogram in the results produced from the simulated data is reweighted using a scaling factor w_{MC}, given by

$$w_{MC} = \frac{\sigma k L}{N} \tag{3.1}$$

where σ is the theoretical cross section, L the integrated luminosity of the real dataset, N the total number of simulated events ($\Sigma_N w_i$) and k the Real K-Factor, which is a correction to the leading order cross section to reproduce the higher order calculation for the interaction. This reweighting of the Monte-Carlo datasets allows valid comparison of the Monte-Carlo simulated events with varying data sample sizes, as in the case of the reduced data luminosity in this analysis.

3.4 Jet Extraction

The analysis is based on the jet objects from the detector contained in the DxAOD, the reconstruction of which is covered in Section **??**. Both the offline jet objects and the online equivalents are retrieved, however the method by which the full collection of jets is assembled differs in either case. For offline jet objects, the DxAOD contains a complete set of jets for each reconstruction algorithm, which are each associated to the relevant jet b-tagging information. Offline jets were calibrated in line with the 20.7 recommendations (Table A.2). In addition, recorded individual jets were required to have $p_{\mathrm{T}} > 45$GeV.

Recovering the trigger-level jet objects from the xAOD is done by assembling distinct object collections from the data into a single jet collection. The jets that satisfied the trigger requirements are stored as *split*-jets in the data. These *split*-jet objects are those that will have b-tagging decisions associated with them. Any duplicate *split*-jets in this collection, determined by pairing jets and removing those with ΔR spacings below a threshold value of 0.3, are removed and the b-tagging information stored in a separate xAOD container is associated with the *split*-jets. Following this all HLT trigger jets are retrieved, which do not possess b-tagging information. The full set of HLT jets is compared to the *split*-jets and any duplicates are removed, again using ΔR matching, from the HLT jet set to form the *nonsplit*-jets. The combination of the *split*-jets and *nonsplit*-jets forms the complete jet collection for the trigger level event.

This complete collection of *split* and *nonsplit*-jets was taken as the comprehensive full set of jets for an event. The lack of associated b-tagging information for the *nonsplit*-jets meant only jets from the *split*-jet list could be designated as b-jets.

3.4.1 *b*-jets

The details of b-jet identification are covered in Section **??**. Offline b-jets were tagged using the *MV2c10*-tagger configured using the January 2017 recommendations (Table A.2) with two defined efficiency working points: *tight*, with an overall efficiency of 70% and *loose* with 85% tagging efficiency. Online b-jets were tagged using the *MV2c20*-tagger as configured during the data taking, which made use of the March 2016 Recommendations (Table A.2) with two identically defined *tight* and *loose* working points. The use of the older b-tagging algorithm in the decision for the online jets was forced. The online trigger objects in the DxAOD only contain the results of the evaluated b-tagging, the quantities used during the calculation were previously discarded. As such, this analysis was required to use the MV2c20 algorithm for HLT jets.

3.5 VBF $H \to b\bar{b}$ Analysis Strategy

Candidate VBF $H \to b\bar{b}$ events are selected by requiring two *central* b-jets which form the Higgs candidate and two high p_{T} VBF-tagging jets. The analysis presented in this disseration follows the selection criteria outlined in the 2016 Search for VBF $H \to b\bar{b}$ at ATLAS [60]. Searches using VBF $H \to b\bar{b}$ consider two exclusive analysis channels of interesting events: the *four-central* channel, which requires all four jets to be contained within the central region $|\eta| < 2.8$, and the *two-central* channel which requires two jets in the central region and at least one forward jet. The analysis presented in this dissertation focuses on the *two-central* channel as this is a more standard VBF topology.

For the *two-central* channel, the event was required to pass the `HLT_j80_bmv2c2070_split_-j60_bmv2c2085_split_j45_320eta490` trigger. This trigger requires a single L1 jet ROI of $E_{\mathrm{T}} > 40\mathrm{GeV}$ and $|\eta| < 2.5$, a second central jet ROI with $E_{\mathrm{T}} > 25$, and a forward jet ROI with $E_{\mathrm{T}} > 20\mathrm{GeV}$ and $3.1 < |\eta| < 4.9$. At the HLT, one central jet b-tagged at the *tight* working point with $p_{\mathrm{T}} > 80\mathrm{GeV}$, and a jet with $p_{\mathrm{T}} > 60\mathrm{GeV}$ tagged at the *loose* working point were both required. Finally a HLT forward jet with $E_{\mathrm{t}} > 45$ between $3.2 < |\eta| < 4.9$ was needed.

Once the trigger was passed, the event was required to contain one jet with $p_{\mathrm{T}} > 95\mathrm{GeV}$ which was b-tagged at the *tight* working point and one additional jet with $p_{\mathrm{T}} > 70\mathrm{GeV}$ that passed the *loose* b-tagging working point. One forward jet with $3.2 < |\eta| < 4.4$

and $p_{\mathrm{T}} > 60$GeV was required along with a final VBF jet with $p_{\mathrm{T}} > 20$GeV and $|\eta| < 4.4$. Finally the p_{T} of the $b\bar{b}$ pair was required to exceed 160 GeV. This cut is to remove kinematic sculpting of the M_{bb} distribution, which for absent or lower $p_{\mathrm{T}bb}$ cuts has a pronounced bump in the 200-300GeV M_{bb} region. This bump is a result of the correlation between m_{bb} and $p_{\mathrm{T}bb}$. By requiring the $p_{\mathrm{T}bb}$ cut the m_{bb} distribution forms a regular falling distribution.

The events were required to be clean events, unaffected by any small detector issues, and the jets were assigned to components of the VBF $H \rightarrow b\bar{b}$ event as described in the following procedure. All pairs of jets that passed the *loose* working point (where either of the jet pair passed the *tight* working point) were considered; the pair with the highest $p_{\mathrm{T}bb}$ was selected as the Higgs candidate. An identical iterative procedure was carried out to assign the VBF pair, using jets not marked for consideration as the Higgs candidate. One of the VBF jet pair was required to satisfy the forward jet selection criterion, and the highest invariant mass pair was selected.

CUT TABLE

These conditions were identical for both the Monte-Carlo simulation and data, with the exception of the trigger requirements which were not required for the simulated samples.

In a full ATLAS analysis [60], the signal is extracted from the selected events using a Boosted Decision Tree trained to extract the VBF $H \rightarrow b\bar{b}$ events over non-Higgs backgrounds. Time constraints in this analysis prohibited a full BDT analysis, but discussion of boosted decision trees and training is covered in Appendix B.

Chapter 4

Object Performance

Prior to conducting a full study of TLA on the VBF $H \to b\bar{b}$ channel, the features of jet objects reconstructed offline and within the HLT were compared to identify any performance differences in the base components of event reconstruction. The jet objects were compared on a one to one basis, by matching an online jet to an offline jet by requiring the ΔR (Section **??**) value between the two jets to be below a threshold value of 0.3. This cut was determined from a plot of ΔR values between all pairs of jets, shown in Figure **??**.

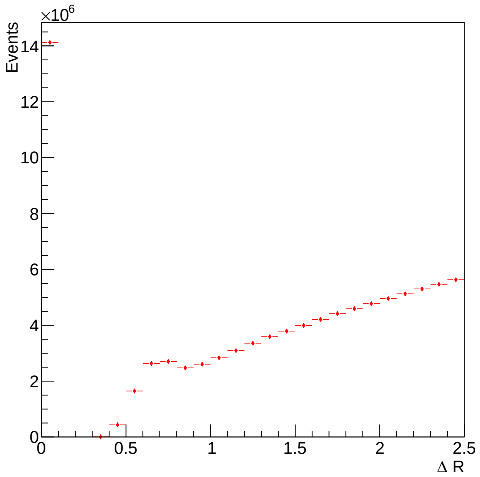

Figure 4.1: Plot of ΔR values for all online/offline jet pairs taken from the Monte-Carlo data. The large spike at ~ 0 accounts for matching jets, with the higher ΔR Values corresponding to differing jet pairs.

To compare the online and offline jets, the ratio of the difference in value for a variety of jet kinematic properties between the matched jets were evaluated. These values were calculated for jet feature X using the ratio of the difference between the offline and online jet features to the offline jet feature

$$\frac{\Delta X}{X} = \frac{X_{Offline} - X_{Online}}{X_{Offline}} \tag{4.1}$$

where $X_{Offline}$ is the value of the kinematic quantity for the offline jet and X_{Online} is the same quantity for the HLT jet. Of the kinematic jet quantities, jet p_{T} was the most significant value to study for a VBF $H \to b\bar{b}$ analysis. In addition, the jet η and ϕ values were compared to assess how similar the topological distribution of the HLT and offline jets was.

These key kinematic quantities were studied for both the leading b-jet and the leading non b-jet for an event given these jet types make up a VBF $H \to b\bar{b}$ event. The jet objects were also divided into buckets of pseudorapidity described in Table **??** in order to examine any changes in behaviour in η, as any differences will significantly impact any assessment of the forward VBF $H \to b\bar{b}$ jets.

Table 4.1: Pseudorapidity bands.

Jet Designation	η Range
Central	$0 < \lvert\eta\rvert < 1$
	$1 < \lvert\eta\rvert < 2.4$
Forward	$2.4 < \lvert\eta\rvert < 4.9$

The jets used to produce these plots were taken from all analysed Monte-Carlo events and all real data events where the `HLT_j80_bmv2c2070_split_j60_bmv2c2085_split_j45_320eta490` trigger was passed, but the additional VBF $H \to b\bar{b}$ requirements mentioned in Section **??** were not enforced. In addition, given the p_{T} requirements of the desired event are high, only jets with $p_{\mathrm{T}} > 45$GeV were considered for analysis.

4.1 Leading *b*-jets

The leading p_{T} offline b-jet was selected from an event, requiring the jet to pass the *Tight* b-tagging working point. This jet was matched to a corresponding online jet using ΔR matching, and the properties of each of these jets compared in both data and Monte-Carlo. The $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ distribution for both the Monte-Carlo and data with respect to the p_{T} of the leading b-jet is shown in Figure **??**.

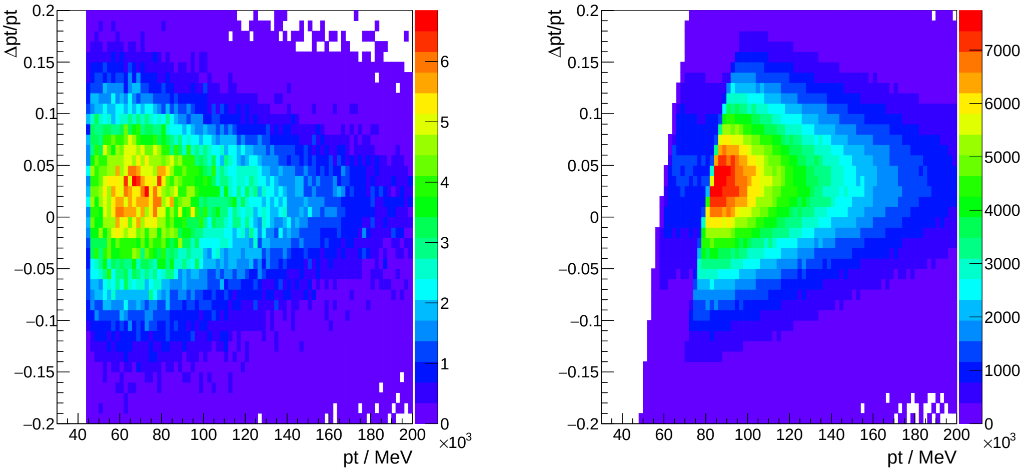

Figure 4.2: $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ for the leading p_{T} b-jet against p_{T} of the offline b-jet, plotted for Monte-Carlo simulation in the left panel and data in the right panel.

The comparative performance of the online and offline jets is p_{T} is broadly similar for events in both data and Monte-Carlo. The bulk of the results occur with a $0 < \frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}} < 0.05$ and the two plots show a comparable distribution drop off, both showing a maximum $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ width of $-0.1 < \frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}} < 0.15$ and showing the p_{T} distribution reaching a maximum of $\sim$ 80GeV. The distinctive curved edge starting at $p_{\mathrm{T}} \sim$ 80GeV present in the real data is the result of the trigger being applied to each event, which was not applied in the Monte-Carlo simulation. The trigger requires at least one jet with a $p_{\mathrm{T}} >$ 80GeV which results in the small number of events below this cut value.

The curve of the distribution shown in the right panel of Figure **??** can be explained given $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ is predominantly positive. In the average case based on this, the p_{T} of the offline jet is higher than the online jet. As the trigger is evaluated on the online jet, only events with an online $p_{\mathrm{T}} >$ 80GeV will be entered into this histogram. For an offline jet with p_{T} = 85GeV to have $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}} = 0.1$, the online jet would be less than the trigger p_{T} cut and as such will not enter into the plot shown in Figure **??**. This exclusion of certain $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ values for certain offline p_{T} values follows from the demonstrated bias in $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$, and produces the curved edge of the distribution.

The distribution of the $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ about 0 can be shown in more detail by taking a slice across the distribution for a representative p_{T} value, which is shown in Figure **??** for leading b-jets with 89 $<p_{\mathrm{T}} <$ 91GeV. The $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ values were also split into the η bands from Table **??**. For the leading b-jet, this is constrained to be within the region of the detector where b-tagging is available, so the forward band is excluded.

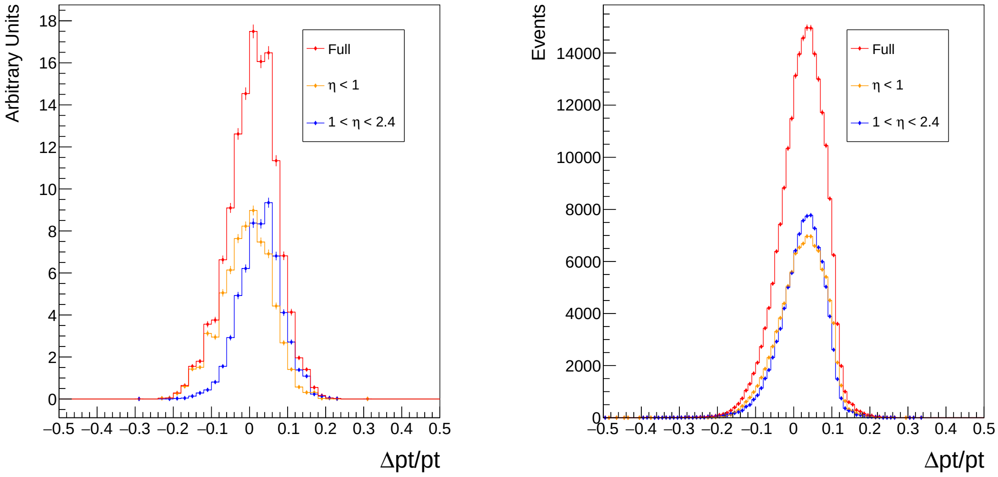

Figure 4.3: $\frac{\Delta p_\mathrm{T}}{p_\mathrm{T}}$ distribution for the leading b-jet with 89 $< p_\mathrm{T} <$ 91 GeV. The distributions for all events and events split by η region are shown. Monte-Carlo simulation is shown in the left panel and data in the right panel.

The results show similar profiles between the Monte-Carlo and Data events for $\frac{\Delta p_\mathrm{T}}{p_\mathrm{T}}$. Both plots show the median offline p_T values to be higher than the online, with a median shift of 4% in Data and 2% in Monte-Carlo. The performance between η ranges was also consistent. The profiles broadly match the full shape of each other, but the Monte-Carlo plot in the left panel of Figure **??** showed a slight difference in $\frac{\Delta p_\mathrm{T}}{p_\mathrm{T}}$ value as the central η range peaked at ~ 0. The breadth of these distributions is quite large, with both Data and Monte-Carlo showing a spread of 10% in $\frac{\Delta p_\mathrm{T}}{p_\mathrm{T}}$.

This offset of the median $\frac{\Delta p_\mathrm{T}}{p_\mathrm{T}}$ value shows that there is a difference in the jet energy calibration between the HLT and the offline reconstruction. The difference between the two is also shown by the offset peaks of the η bands in Figure **??**, with the more central region performing better. Prior calibration studies of the ATLAS calorimeter have shown the energy readouts to be more consistent towards the central regions of the detector [56]. This could cause the inaccuracy of the trigger jets in the higher pseudorapidity regions as offline jet reconstruction can make use of developed calibration tools to account for these differences. Using these standard tools, the energy scale calibration difference between the offline and online jets can be rectified for future analyses [56, 57].

The $\frac{\Delta X}{X}$ comparisons can be carried out for the topological jet properties (η, ϕ) to confirm the offline and online jets are positioned within the detector in a similar fashion. Plots of $\frac{\Delta\eta}{\eta}$ against the pseudorapidity of the offline jet in the selected pair for data and Monte-Carlo simulation are shown in Figure **??**, and comparable plots of $\frac{\Delta\phi}{\phi}$ against the offline ϕ are given in Figure **??**.

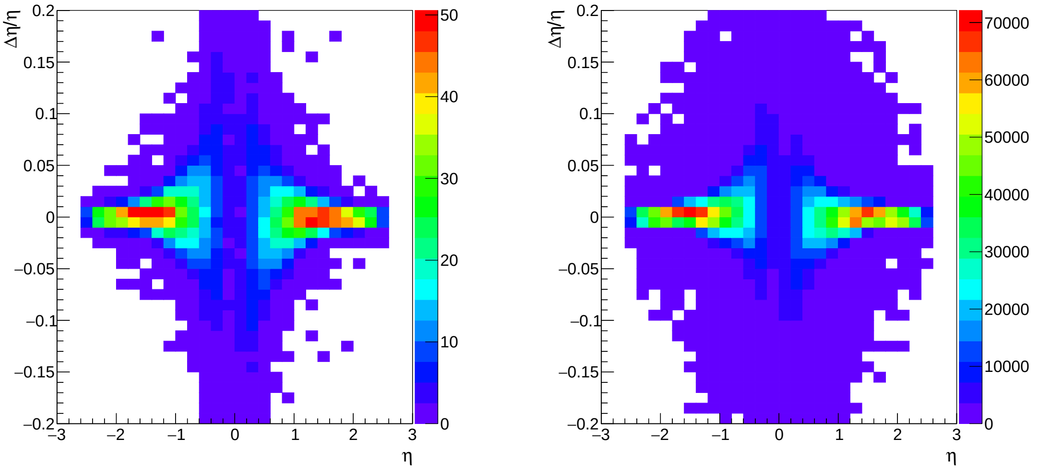

Figure 4.4: $\frac{\Delta\eta}{\eta}$ for the leading b-jet, for Monte-Carlo simulation in the left panel and data in the right panel.

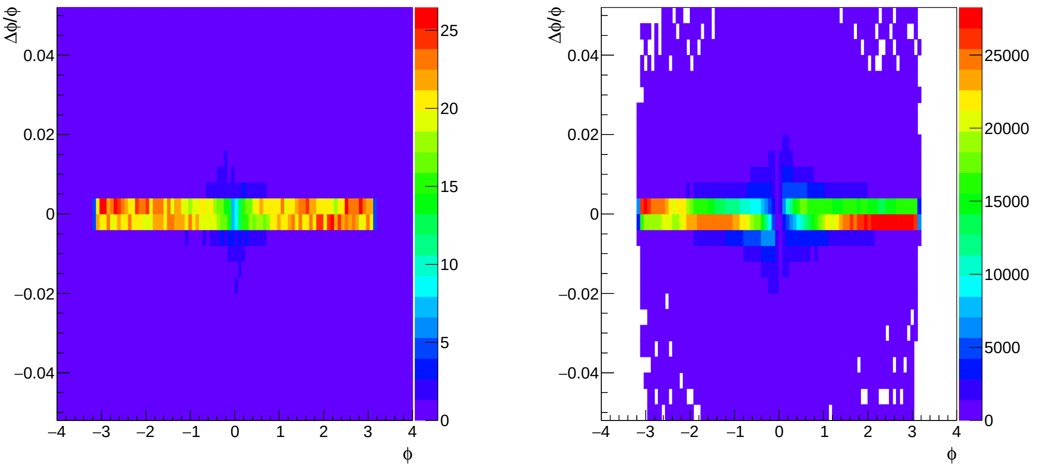

Figure 4.5: $\frac{\Delta\phi}{\phi}$ for the leading b-jet, for Monte-Carlo simulation in the left panel and data in the right panel.

The data and Monte-Carlo distributions for these values are extremely similar to each other, and also show very close agreement between the values for offline and online jet objects. For both $\frac{\Delta\eta}{\eta}$ and $\frac{\Delta\phi}{\phi}$ the median value is ~ 0 and the width of the distribution is less than 1% of the value. These results show the (η, ϕ) positions of the online and offline jet objects are comparable to each other.

4.2 Leading Non b-jets

For VBF $H \longrightarrow b\bar{b}$, a pair of high p_{T} forward jets is the other significant feature, so the offline/online performance in the leading non b-jet was studied. Identically to the analysis of the leading b-jet in Section **??**, the p_{T} , η and ϕ values of a matched offline/online jet pair were studied by calculating $\frac{\Delta X}{X}$ values and plotting against the offline kinematic quantity. The results could be split into the η bands from Table **??**, with the forward pseudorapidity band available for analysis as b-tagging was not required. Plots of $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ for the leading non b-jet are shown in Figure **??**.

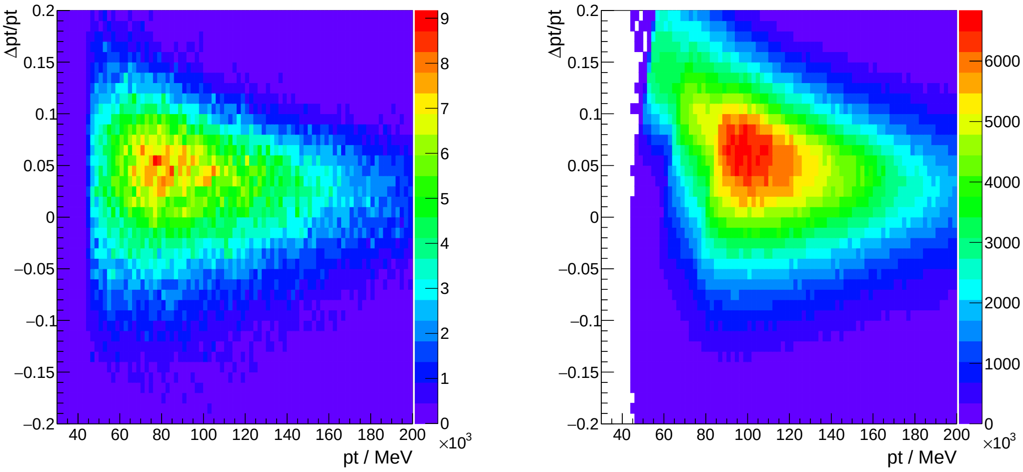

Figure 4.6: $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ for the leading p_{T} non b-jet against p_{T} of the offline jet, plotted for Monte-Carlo simulation in the left panel and real data in the right.

The leading non b-jet distributions show similar results to the leading b-jetdistributions in Figure **??**. The peak of the distribution between $0 < \frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}} < 0.1$ shows there is agreement between the p_{T} of the offline and the online non b-jet. The overall shape of the distribution shows some differences between the Monte-Carlo simulation and data however. The distributions are similarly structured, with a $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ width between -0.1 and 0.15 and the p_{T} offline distribution reaching a maximum value of $\sim$ 180GeV. However, there is a distinct cluster of results shown only in the right panel of Figure **??** of low p_{T} offline jets with $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}} > 0.1$. There is also a suggestion of a curving edge to the distribution for the data, in an opposite direction to that shown for the leading b-jet in Figure **??**. In addition, the peak of the data is slightly higher in p_{T} ($\sim$ 80-120GeV) than in the Monte-Carlo ($\sim$ 60-110GeV).

The slight upward shift in p_{T} can be explained by the p_{T} requirements of the trigger applied only to the data. Requiring the jet components to exceed high p_{T} cuts will bias the results to events containing high p_{T} jets, accounting for the upward p_{T} shift of the data events in the right panel of Figure **??** relative to the left.

As for the leading b-jet , slices can be taken of the $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ distribution to show the spread of values more clearly. Plots of $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ values for leading non b-jets with $89 < p_{\mathrm{T}} < 91$GeV are shown for Monte-Carlo simulation and data in Figure **??**, and results have been split into the pseudorapidity bands from Table **??**.

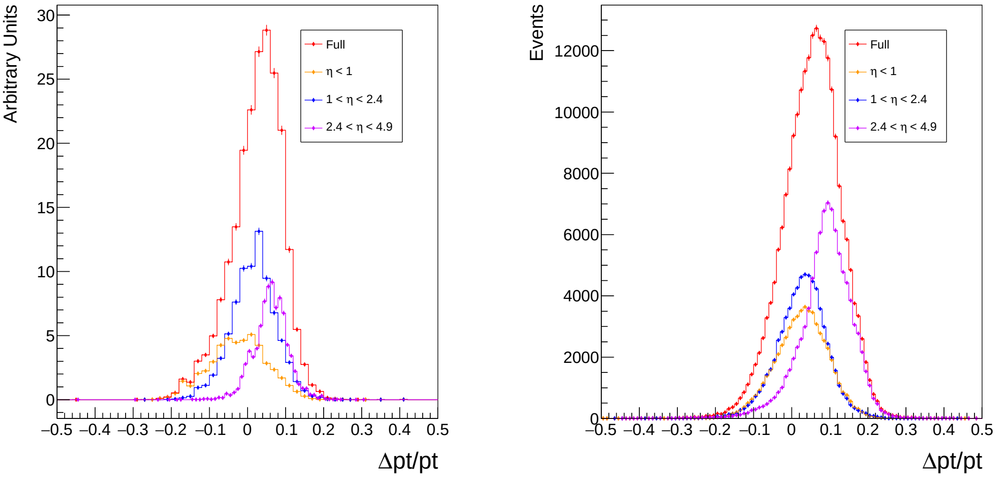

Figure 4.7: $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ distribution for the leading non b-jet with $89 < p_{\mathrm{T}} < 91$ GeV plotted for Monte-Carlo simulation in the left panel and data in the right panel. The distributions for all events and events split by η region are shown.

Both Monte-Carlo simulations and data show the median value for offline jet p_{T} to be higher than the online jet by 4% and 6% respectively. The overall distribution shape is similar between the simulated and real events for the full set of results, but the distributions for the η bands differ between the Monte-Carlo and the real data.

The Monte-Carlo results for the central η band show a dip in p_{T} at the centre of the distribution and are shifted in $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ towards the negative. Both the data and Monte-Carlo show that the $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ value is much closer to 0 for the two central η bands than the forward band, which peaks significantly higher than the median $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ value. The offset of the forward η band from the median is much worse for the dat in the right panel of Figure **??**. In addition, the relative proportions of the three η bands differ. In Monte-Carlo results most jets fell in the middle $1 < |\eta| < 2.4$ while data showed significantly more forward jets.

The relatively increased proportion of forward jets is likely a consequence of the `HLT_j80_bmv2c2070_split_j60_bmv2c2085_split_j45_320eta490` trigger being applied to the data. As the data events are required to have a forward jet to be stored in the histogram, this will bias the results to contain a greater proportion of forward jets, leading to the larger peak.

The $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ results for the leading non b-jet as for the leading b-jetshow a difference in energy calibration between the HLT jet objects and the reconstructed offline objects. This difference in calibration can be corrected using standard jet calibration tools to bring the p_{T} values into closer agreement with one another [56, 57].

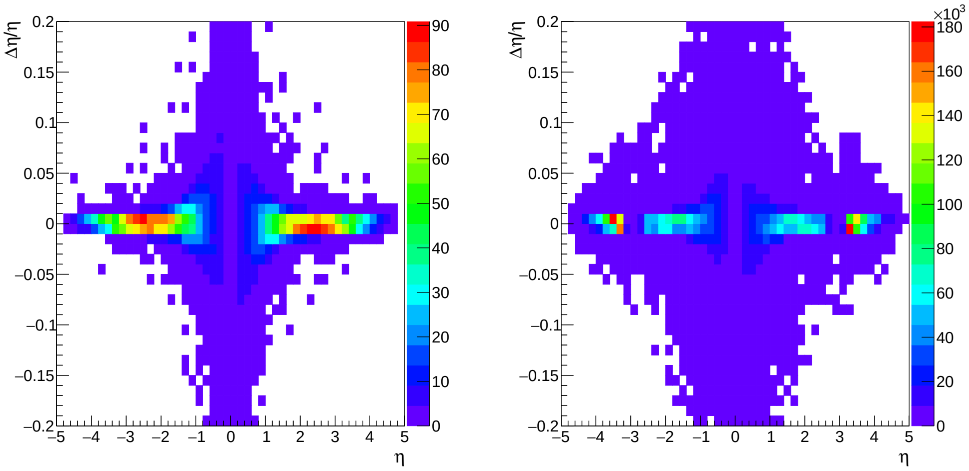

Figure 4.8: $\frac{\Delta\eta}{\eta}$ for the leading non b-jet, for Monte-Carlo simulation in the left panel and data in the right panel.

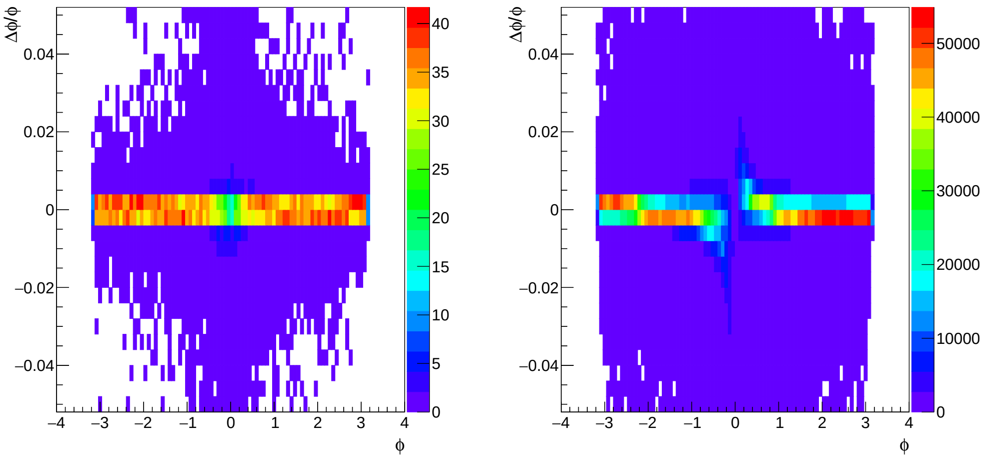

Figure 4.9: $\frac{\Delta\phi}{\phi}$ for the leading non b-jet, for Monte-Carlo simulation in the left panel and data in the right panel.

These $\frac{\Delta X}{X}$ values can be calculated and plotted for the topological kinematic quantities (η, ϕ), with $\frac{\Delta\eta}{\eta}$ for the leading non b-jet plotted in Figure **??** against the offline jet η, and $\frac{\Delta\phi}{\phi}$ against offline ϕ plotted in Figure **??**. As with the b-jets the (η, ϕ) values of the offline and online jets produce nearly identical results, with the distribution of $\frac{\Delta X}{X}$ firmly centred around 0 and a width of less than 1%. As with the b-jets this shows the spatial position of the leading non b-jet is comparable for online and offline objects.

4.2.1 Summary of Comparison of Jet Objects between Offline and Online

The jet objects reconstructed in the HLT have some slight differences in the reported values for key topological variables, but overall they perform in a similar fashion, both in Monte-Carlo simulations and in Real data. The positional variables, ϕ and η are directly comparable between offline and online jet objects, with the majority of objects having values with $< 1\%$ disagreement for both b-jets and non b-jets . For the p_{T} of jet objects, the values are not in perfect agreement, but have a consistent offset observed in Monte-Carlo simulation and data.

This difference in jet energy scale calibration can easily be overcome by constructing specific jet calibrations using already standard jet calibration tools [56, 57] to correct the offset of the p_{T} values.

With this calibration executed on the HLT jet objects, the online jets would then be directly comparable in energy scale and topographical location to the offline jet objects, and as such would be usable in analyses as a replacement for the offline objects. Further verification of this could be carried out by emulating the trigger for the Monte-Carlo simulation to check if the same features arise in the kinematic quantity distributions.

4.3 Jet Tagging Efficiency

As covered in Section **??**, the standard algorithm for 2016 physics analyses was chosen to be the 2016 MV2c10 algorithm. However, the HLT b-tagging algorithm uses the older MV2c20 algorithm [4]. In order for any form a Trigger Level Analysis to be considered valid, the performance of the tagging algorithms used in the trigger, which are fixed at the point of data collection, must be comparable with the tagging executed offline with more up to date b-tagging configurations.

To study this, the b-tagging efficiency at trigger level and offline is studied for different jet flavours using the MC sample. The Monte-Carlo sample was used as the *truth* nature of the jet object is known, and the result of the b-tagging algorithm can be comparedfor b-jets c-jets and light-jets.

In the analysis, an offline/HLT jet pair was formed using $\mathit{\Delta R}$ matching and truth label of the offline jet used to assign a flavour to the pair. Light-jets, b-jets and c-jets were all studied separately to view the b-tagging efficiency and the mistag rate of both algorithms operating at the *tight* for an expected b-tagging efficiency of 70%. The efficiency plots in Figures **??**, **??** and **??** show the fraction of these jets that were identified as b-jets by the HLT and offline tagging algorithms. These plots were created from events following the same cuts as for the leading b-jet and non b-jet as discussed at the beginning of the chapter.

4.3.1 *b*-jet efficiency

For jets labeled as true b-jets, the tagging efficiency of can be calculated and plotted against kinematic quantities of the b-jets. Figure **??** shows the b-tagging efficiency ϵ plotted against the p_{T} and η of the offline b-jet.

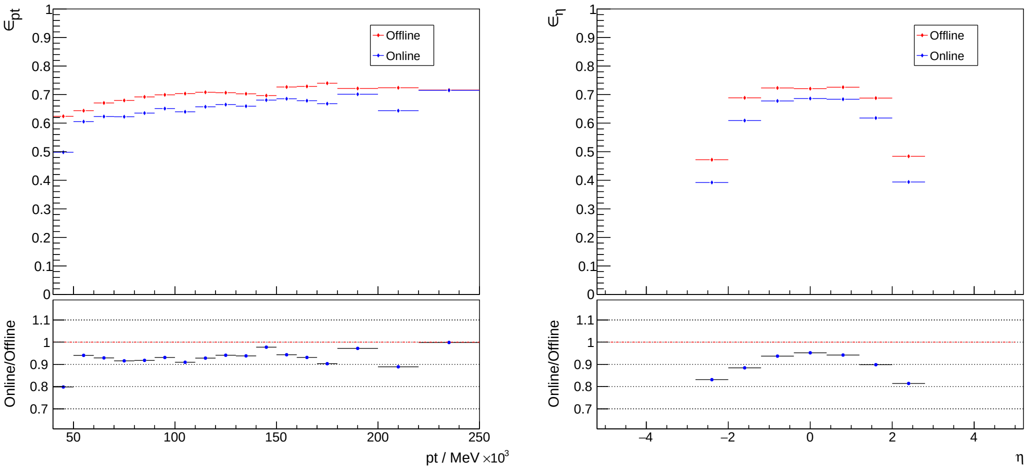

Figure 4.10: b-tagging efficiency for truth b-jets in Monte-Carlo simulation, evaluated for offline jets with the 2016 MV2c10 algorithm and for online jets with the 2015 MV2c20 algorithm, plotted against offline jet p_{T} in the left panel and offline jet η in the right panel.

The overall distribution shape in p_{T} and η for the b-tagging efficiency is consistent for both the online and offline b-jetṪhe shape of the distribution in p_{T} shown in the left panel of Figure **??** is consistent with the efficiency curves expected for the MV2 b-tagging algorithm with respect to p_{T} [46], and the efficiency is consistent the 70% value expected for the *tight* working point for the offline jets, shown clearly by the flat peak of the η distribution in the right panel for the central η regions where the p_{T} distribution of jets should be unbiased.

However, the HLT b-tagging is shown to be around 5% less efficient than the offline b-tagging for jets with $p_{\mathrm{T}} > 50$GeV. This value is consistent across the p_{T} distribution shown by the flat line at ~ 0.95 in the ratio plot in the left panel of Figure **??**. The improvement in efficiency between the 2016 MV2c10 and 2015 MV2c20 algorithms is consistent with the comparative behaviour shown for a training $t\bar{t}$ sample [46], but of a larger magnitude.

This change in efficiency, consistent with the differences in the two algorithms, could be rectified by applying the newer algorithm to the online jet objects. However, the trigger-level jet objects in the xAOD sample data only contained the discriminant values from the applied 2015 MV2c20 algorithm which could be extracted using standard b-tagging tools in AnalysisBase. The input variables used for the training and evaluation of the component algorithms of the MV2 algorithms (Section **??**) were discarded from the HLT level jet objects. Retaining this quantities on the online jets would allow future trigger-level analyses to make use of the newer b-tagging algorithms or retrain one algorithm to increase the performance to offline levels. This would result in an increased detector readout size in bytes, which could reduce the permitted rate increase for a TLA. An estimate of such a resultant decrease in rate is not made in this thesis.

4.3.2 *c*-jet efficiency

The same efficiency plot could be produced for c-jets against the kinematic jet quantities. Any result marked as a b-jet for a truth c-jet is a mistagged jet, and plots of the efficiency show measurements of the mistag rate of the algorithm. The mistag rate is plotted in Figure **??** for the online and offline jets against offline p_{T} and η.

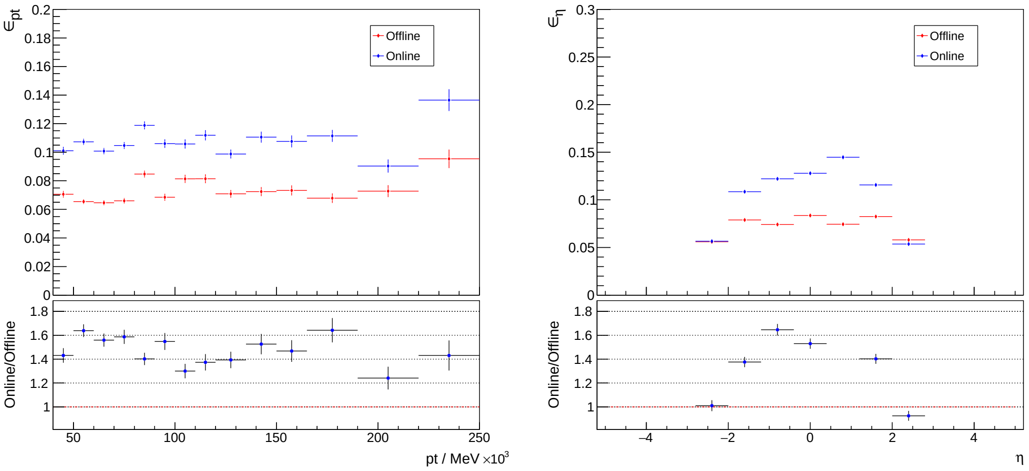

Figure 4.11: Mistag rate for truth c-jets in Monte-Carlo simulation, evaluated for offline jets with the 2016 MV2c10 algorithm and for online jets with the 2015 MV2c20 algorithm, plotted against offline jet p_{T} in the left panel and offline jet η in the right panel.

The shape of the mistag rate distribution is more noisy than the b-jet efficiency plots in Figure **??**, but across the p_{T} distribution in the left panel of Figure **??** the online mistag rate is ∼ 50% higher than the offline rate.

The increase in the rate of c-jet mistagging is absolutely consistent with the refinements to the algorithm between the 2016 MV2c10 and 2015 MV2c20, with increased levels of c-jet rejection in the offline 2016 MV2c10, and the increase is consistent with the expected shift from the optimised algorithm [46]. Similar to the solution for changes in b-tagging efficiency in Section **??**, retaining the inputs to the b-tagging algorithms would allow improved versions of the b-tagging algorithms to be applied to the HLT jets.

4.3.3 Light-jet efficiency

Plots of the mistag rate for truth light-jets for online and offline b-tagging against the offline jet p_T and η are shown in Figure **??**

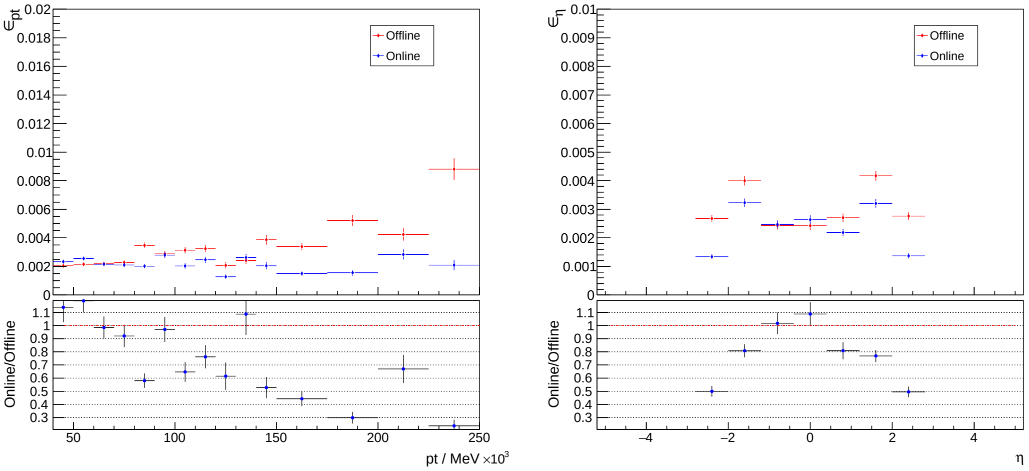

Figure 4.12: Mistag rate for truth light-jets in Monte-Carlo simulation, evaluated for offline jets with the 2016 MV2c10 algorithm and for online jets with the 2015 MV2c20 algorithm, plotted against offline jet p_T in the left panel and offline jet η in the right panel

The light-jet efficiency plots are noisier than the plots for truth b-jets and c-jets owing the the low mistag rate of ∼ 0.3% shown in the left panel of Figure **??**. For these light-jets, the online algorithm performs better than the offline algorithm overall, with a mistag rate of ∼ 80% the offline rate for jets with $p_T < 150$. The change in the performance for the light-jet rejection, with the 2015 MV2c20 algorithm performing better for higher p_T values and worse for lower p_T values is consistent with the expected change in behaviour between the two algorithms [46].

4.3.4 Tag Matching

For each pair of jets that could be matched between online and offline, and then successfully have a b-tagging decision evaluated on the jets, the agreement of the b-tagging between the two jets was checked. These were found to match one another in 91% of cases.

4.4 Summary

The aim of this section of the analysis was to show that using online reconstructions of the constituent jet objects used in a VBF $H \to b\bar{b}$ analysis was comparable to using offline objects by showing the properties of the jets and the b-tagging of the jets to be similar. The overall performance using the HLT objects as constructed during the data-taking and in the Monte-Carlo simulations is similar to the offline behaviour. The topological jet quantities (η, ϕ) are directly comparable between the two types of jet objects. However there are differences between the p_{T} values of the HLT and offline jet objects and differences between the performances of the b-tagging algorithms.

These differences are readily rectifiable however. The energy scale calibration differences can be accounted for using standard jet calibration tools to bring the p_{T} values of the online and offline jets into agreement with one another [56], while the b-tagging performance can be made similar if the input variables to the MV2 algorithm are preserved on the trigger-level jet object, such that more developed b-tagging algorithms can be applied to the jet instead to the algorithm used during data collection.

With these corrections, there are no differences between the trigger level objects making up a VBF $H \to b\bar{b}$ event and the offline objects that would prohibit a TLA analysis of the VBF $H \to b\bar{b}$ channel.

CHAPTER 5

VBF $H \rightarrow b\bar{b}$ Analysis

After comparing the base constituents of the VBF $H \rightarrow b\bar{b}$ event between the offline and HLT level and finding them to be similar in behaviour, the specific objects that make up a VBF $H \rightarrow b\bar{b}$ event can be studied and compared. In this section, the events were required to pass all cuts discussed in Section **??** and the designation of the jets as b_i, j_i is highlighted in that section.

5.1 Cutflow

Prior to investigating the core kinematic variables and the more complex kinematic variables used for the Boosted Decision Tree training (Appendix B), the event cutflow for both the Monte-Carlo and real data should be studied to highlight any differences between the event counts. The event counts are given in Table **??**, and the ratio of the events is shown in Figures **??** and **??**.

5.1.1 Monte-Carlo

Overall, the online performance has fewer events than the offline for all points in the cutflow, and overall produces $\sim$ 80% of the total signal events. There are three distinct jumps in the cutflow ratio, at the cuts on the *loose* b-jets , light-jets and the forward jet requirement, of $\sim$ 7% each. As shown in Figure **??**, online b-tagging is $\sim$ 93% as efficient as the offline b-tagging. When considering tagging two distinct b-jets , any difference in efficiency is squared. Given the difference in tagging rates, this would result in $\sim$ 86% tagging efficiency

Table 5.1: Cutflow for the *two-central* VBF $H \rightarrow b\bar{b}$ events as described in Section **??**. The cutflows are given for the online and offline channels in both data and Monte-Carlo along with the percentage of original events.

Cut	MC Offline	MC Online	Data Offline	Data Online
Clean Events	6229.48	6229.48	150611000	150611000
Trigger	6229.48	6229.48	6679390	6679390
$\geq$ 2 *loose* b-jets	503.552	467.146	2275760	2932620
$\geq$ 2 light-jets	483.499	417.845	2189700	2671280
Tight b-jet requirement	330.962	288.806	1490320	1640290
Forward jet requirement	51.843	40.8484	1186610	958414
$p_{\mathrm{T}bb} > 160$GeV	32.7426	26.7038	309454	259411

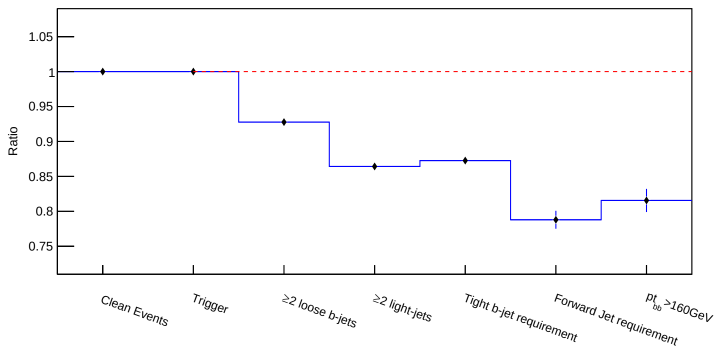

Figure 5.1: Ratio of the online event count over the offline event count for the Monte-Carlo

for two b-jets , which is lower than shown in the cutflow. As shown for the leading b-jetin Section **??**, the offline jet is typically higher in p_{T} than the online jets. However the difference is small , $\sim$ 2%, so any effect on the cutflow should not be as pronounced.

The $\sim$ 7% drop on the light jet requirement is unexpected, the requirement was solely for 2 jets with $p_{\mathrm{T}} > 20$GeV. Given the points above with respect to the p_{T} difference between online and offline, this drop should not be so sever. The fact the p_{T} cuts on the light jets were so low also suggests an anomalous result here as such a cut should not contribute a significant reduction in either online or offline.

Following the drop for the light-jet cut, there is an unexpected increase in the online ratio following the *tight* b-tagging cut. As highlighted, the tagging efficiency was worse for online than offline, so any requirement for a tagged b-jet would be expected to produce a decrease in the online event count relative to the offline count. Perhaps spuriously, the cutflow at this point corresponds to the 86% figure expected given the relative tagging efficiency for two b-jets.

The final drop occurs following the requirement for a high p_{T} forward jet. Figure **??** shows that for non b-jets in the forward region of the detector, the p_{T} of the offline jet is consistently higher than the online p_{T} . This difference would result in a drop in the online events, with fewer jets passing the threshold p_{T} cut compared to the offline events.

Overall for the Monte-Carlo events, there was a 20% reduction in the number of events that passed the VBF $H \longrightarrow b\bar{b}$ cuts.

5.1.2 Data

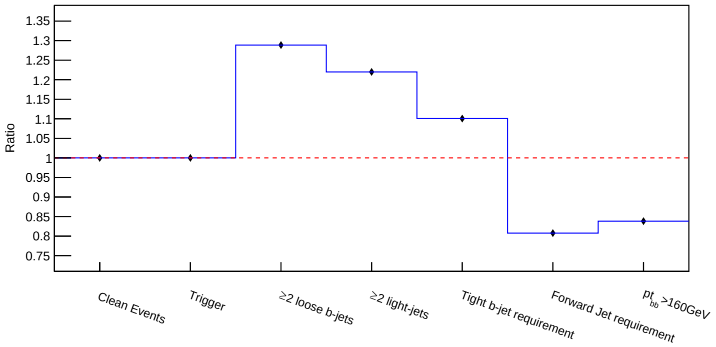

Figure 5.2: Ratio of the online event count over the offline event count for the real data

5.1.3 Summary of Cutflow Comparison for the Online and Offline VBF $H \longrightarrow b\bar{b}$ events

While the granular details of the cutflow (LEAVE A LOT TO BE DESIRED), the overall effect on the reduction of events to the final VBF $H \longrightarrow b\bar{b}$ event state is consistent between both the Monte-Carlo simulations and data.

The final relative reduction in online event count relative to the offline event count is shown for the Monte-Carlo simulations in Figure **??** to be to ~ 82% of the offline event count. Similarly the ratio plot for the data events in Figure **??** results in ~ 84% final state events for the data events. The consistent values for the rate reduction of a trigger-object based analysis suggest that future analyses may be able to apply TLA to the VBF $H \longrightarrow b\bar{b}$ channel to increase the event yield in these analyses and improve the statistical significance of any results.

With an average ~ 83% reduction in event rate for the channel, usage of TLA at a rate of 2kHz as described in Ref. [**?**] and Section **??** would result in an increase in output events of ~ 66% compared to a standard offline analysis.

This statement is an estimate, there is additional computational cost involved with storing and computing the quantities required for the VBF $H \to b\bar{b}$ analysis, such as the increased event size mentioned in Section **??** as a result of storing the b-tagging training quantities to apply the updated algorithms. A rate analysis to ensure the TLA can be applied in the VBF $H \to b\bar{b}$ channel without decreasing the rate increase down to the point the final TLA event count is no longer an improvement is a necessary step before approving TLA, but is beyond the scope of this dissertation.

5.2 Specific Jet Feature Distributions

While the previous chapter showed that the b-jets and non-b-jets had slight differences that could be rectified for future analyses, the behaviour is sufficiently similar that plots of the kinematic quantities of the VBF $H \to b\bar{b}$ jets can be made to ensure behaviour is consistent for the specific jets that make up the final state from Table **??**.

Table 5.2: m_{bb} bins defining an event as signal or background, along with the data source for the quantities.

Designation	m_{bb} range / GeV	Sample
Background (Lower)	m_{bb} <100	Data
Signal	100 <m_{bb} < 140	Monte-Carlo
Background (Upper)	140 <m_{bb}	Data

These plots were presented in signal and background regions, as defined by the m_{bb} value as shown in Table **??**. The signal was plotted only using the Monte-Carlo simulation while the background regions were taken from Data events. The kinematic quantities for jets b_1 and j_1 as designated in Section **??** are plotted for both the online and offline objects. The p_T distributions of b_1 and j_1 are shown in Figures **??** and **??** respectively, while the pseudorapidity of b_1 is shown in Figure **??** and for j_1 in **??**.

The ratio plot in the left panel of Figure **??** shows a flat line in both the upper and lower regions just above 0.8 for the ratio of the number of online jets with respect to the offline jets, consistent with the decrease in final event count shown in Figure **??**. The Monte-Carlo signal plot in the right panel is noisy but shows the ratio to be $\sim$ 0.8 across the p_T range, consistent with Figure **??**. As expected based on the higher m_{bb} requirement the background events for the Upper region are biased towards higher p_T compared to the lower background region.

The relative online and offline behaviour of j_1 is slightly more complicated that jet b_1. Both the data and Monte-Carlo simulation show a consistent curve in behaviour, with online

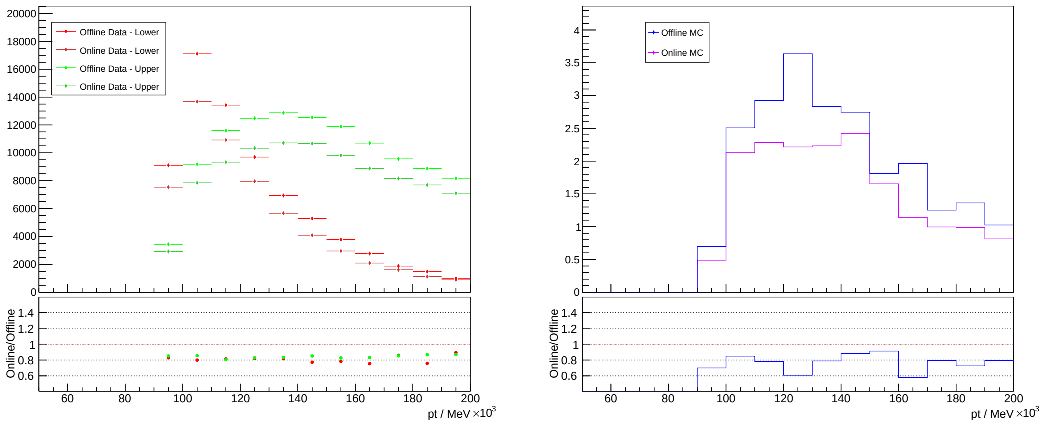

Figure 5.3: p_{T} distribution of the leading b-jet of the VBF $H \longrightarrow b\bar{b}$ event, plotted for both the backgroud data regions in the left panel and Monte-Carlo signal events in the right panel.

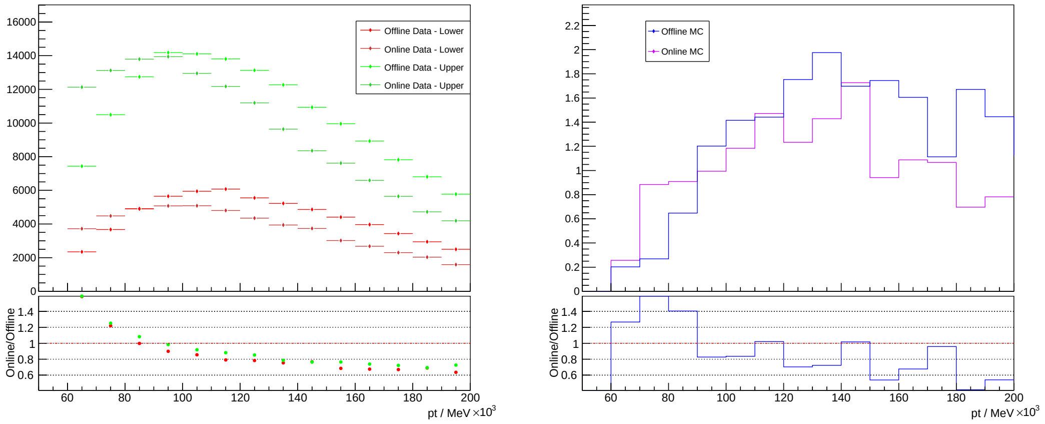

Figure 5.4: p_{T} distribution of the leading non-b-jet of the VBF $H \longrightarrow b\bar{b}$ event, plotted for both the backgroud data regions in the left panel and Monte-Carlo signal events in the right panel.

events being more numerous at low p_{T} than offline, before the ratio tails off below unity at $\sim$ 90GeV in both the plots in Figure **??** to $\sim$ 0.7 in the data. The Monte-Carlo plot jumps around, but the comparitive online/offline performance trends in the same fashion as the background plot.

Figure **??** shows shared distribution shapes between the online and offline objects for both signal and background, and the consistency of the online distribution against the offline is also shown for the leading non-b-jet in Figure **??**. The data distribution for b_1 in the

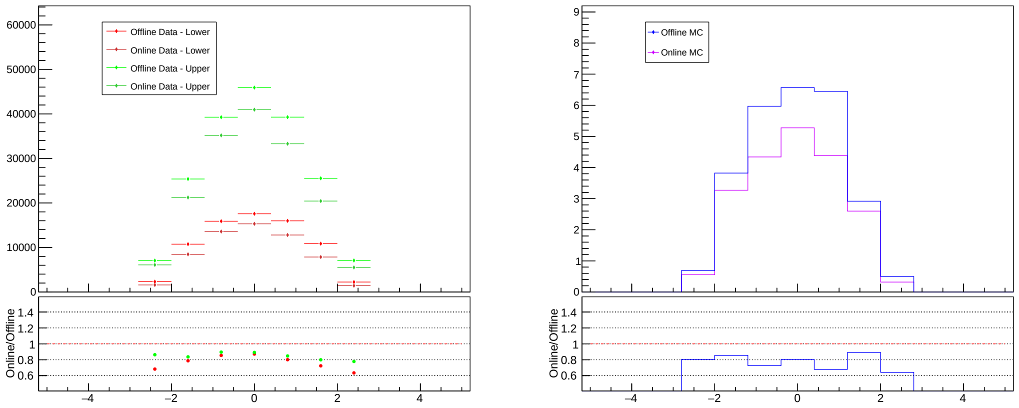

Figure 5.5: η distribution of the leading b-jet of the VBF $H \to b\bar{b}$ event, plotted for both the backgroud data regions in the left panel and Monte-Carlo signal events in the right panel.

left panel of Figure **??** shows a pronounced curve in the online and offline ratio, with more online events surviving the cuts in the central regions, and the overall efficiency being roughly consistent with the expected ~ 0.8 value. The signal distribution in the right panel shows a flatter ratio distribution, also around this value.

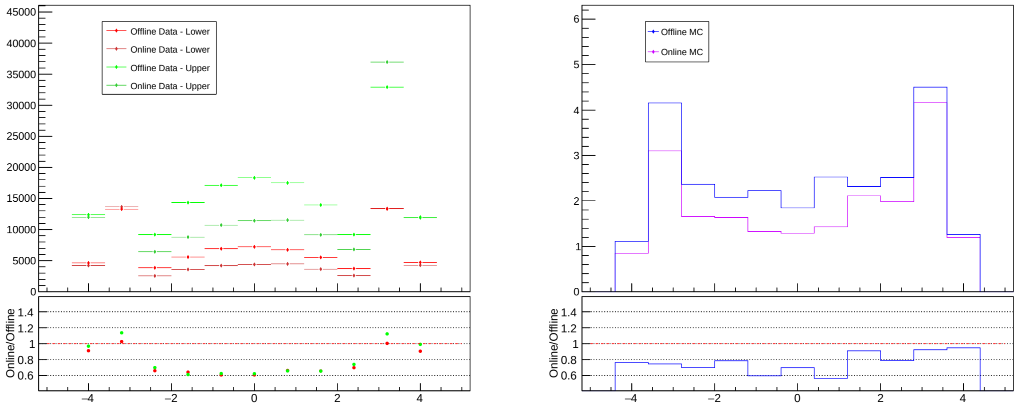

Figure 5.6: η distribution of the leading non-b-jet of the VBF $H \to b\bar{b}$ event, plotted for both the backgroud data regions in the left panel and Monte-Carlo signal events in the right panel.

The η distribution for the leading non-b-jet shows spikes in forward region both the background and signal plots in Figure **??**, which is expected given the jet criteria required for a VBF $H \to b\bar{b}$ event. Both plots show the decrease in the online events relative to the offline events expected in each of these.

5.2.1 Summary of the VBF $H \to b\bar{b}$ jet objects

The plots for the VBF $H \to b\bar{b}$ final state mostly show agreement, with respect to the overall decrease in the event rate for online events, between the online and offline distributions of the kinematic quantities of the jets for signal and background. The shapes if the distributions are consistent between the background and signal events, and the specific features like the forward jet spikes in Figure **??** and the increase p_{T} of the upper background sector in Figure **??** are explainable by considering the topology of a VBF $H \to b\bar{b}$ event.

There are two sections of these plots where the online/offline comparison significantly deviates from the demonstrated efficiency of $\sim 80\%$, in Figure **??** for low p_{T} non-b-jetsand in Figure **??** for the high η non-b-jets. Both these artifacts could be a result of additional forward jets in the online events. A forward jet compared to a jet with identical momentum jet at a central eta will have a lower p_{T} given the geometry of the jet and the definition of p_{T} .

An additional difference was the upper background sector frequently had a higher ratio than the lower background sector, as shown across the entire p_{T} range in Figures **??** and **??**.

5.3 Kinematic quantities of the VBF $H \to b\bar{b}$ event

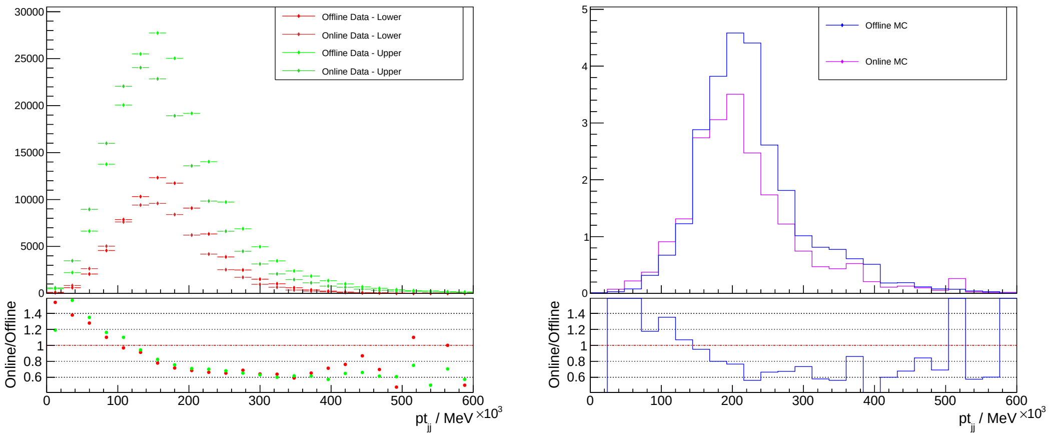

Figure 5.7: $p_{\mathrm{T}jj}$ distribution for the online and offline VBF $H \to b\bar{b}$ events, with background events from data shown in the left panel and Monte-Carlo signal events in the right.

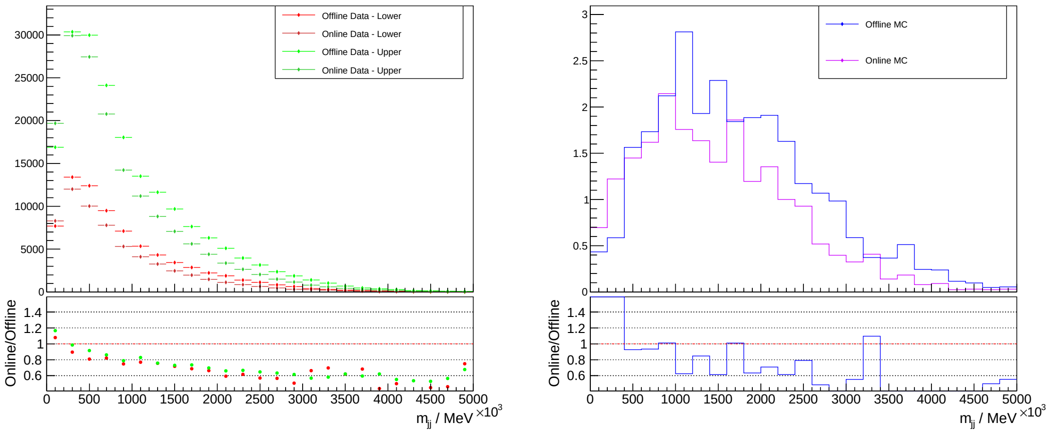

Figure 5.8: m_{jj} distribution for the online and offline VBF $H \to b\bar{b}$ events, with background events from data shown in the left panel and Monte-Carlo signal events in the right.

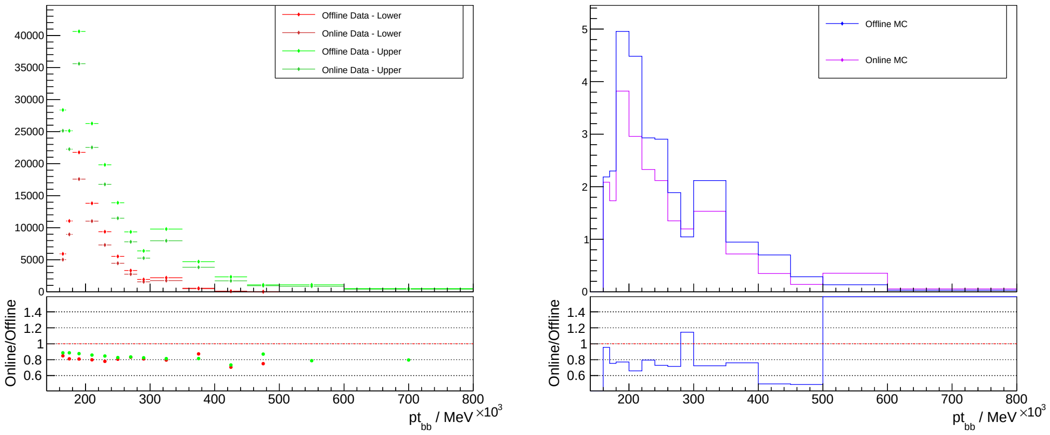

Figure 5.9: $p_{\mathrm{T}bb}$ distribution for the online and offline VBF $H \to b\bar{b}$ events, with background events from data shown in the left panel and Monte-Carlo signal events in the right.

5.4 BDT Input Variables

As discussed in Section **??**, the full BDT analysis of the VBF $H \to b\bar{b}$ search described in Appendix B was not performed for this dissertation. Given this is a critical component of the full Higgs search [60] in VBF $H \to b\bar{b}$, the performance of select BDT variables is explored for the signal and background regions described in Table **??**. The variables m_{jj} and $p_{\mathrm{T}jj}$ covered in the previous section are both BDT training variables.

This section covers η^*, given by

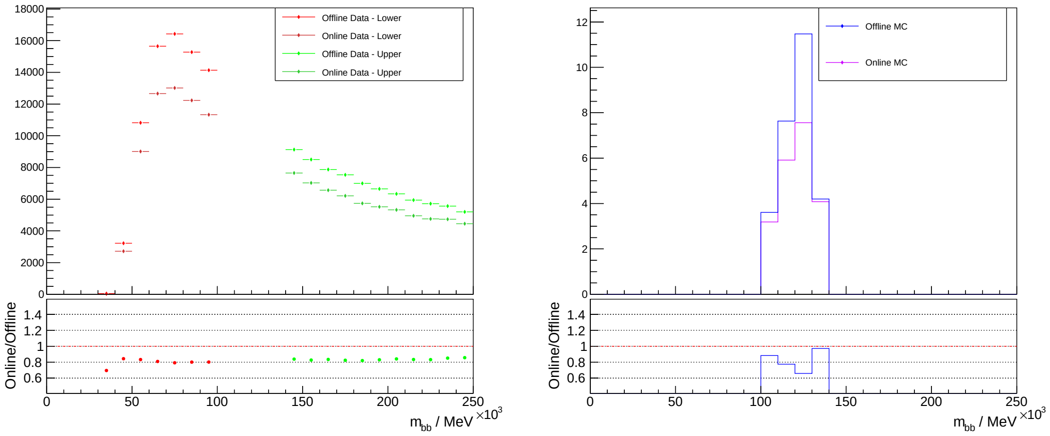

Figure 5.10: m_{bb} distribution for the online and offline VBF $H \rightarrow b\bar{b}$ events, with background events from data shown in the left panel and Monte-Carlo signal events in the right.

$$\eta^* = \frac{1}{2}(|\eta_{j1}| + |\eta_{j2}| - |\eta_{b1}| - |\eta_{b2}|) \tag{5.1}$$

which is plotted in Figure **??** for the online and offline events, and the p_{T} *balance*, given by

$$p_{\mathrm{T}balance} = \frac{\boldsymbol{p_{\mathrm{T}j1} + p_{\mathrm{T}j2} + p_{\mathrm{T}b1} + p_{\mathrm{T}b2}}}{p_{\mathrm{T}j1} + p_{\mathrm{T}j2} + p_{\mathrm{T}b1} + p_{\mathrm{T}b2}} \tag{5.2}$$

which is shown in Figure **??**

As for the VBF $H \rightarrow b\bar{b}$ jet kinematic quantities in Section **??**, these variables behave in the same fashion. For the lower background, upper background and signal sectors the shapes of the online and offline curves are comparable. The relative ratio of the online to the offline events is .8, as covered in Section **??** and shown clearly by the ratio plots in the left panels of Figures **??** and **??**. The plots for the Monte-Carlo signal are noisier, but show a rough 20% decrease in online events compared to offline, and in general the upper background region performs better than the lower region.

Given this consistent behaviour for the BDT quantities derived from the VBF $H \rightarrow b\bar{b}$ event, the trigger-level objects should perform comparably to the offline objects, and as such can be used to train a BDT to refine the analysis into a VBF $H \rightarrow b\bar{b}$ phase space using solely trigger-level objects.

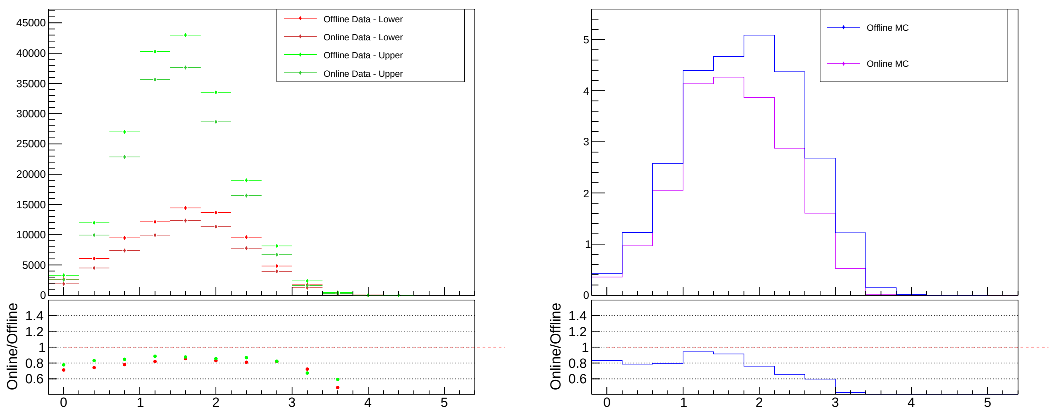

Figure 5.11: η^* distribution for the online and offline VBF $H \to b\bar{b}$ events, with background events from data shown in the left panel and Monte-Carlo signal events in the right.

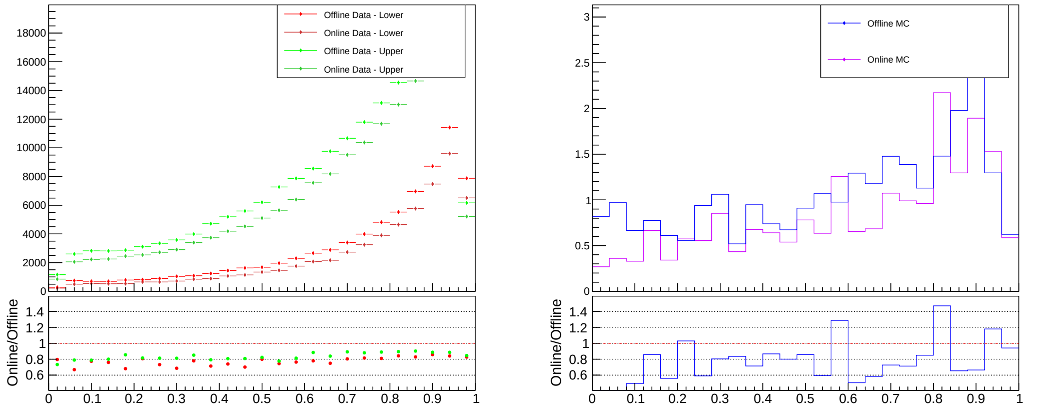

Figure 5.12: $p_{\mathrm{T}balance}$ distribution for the online and offline VBF $H \to b\bar{b}$ events, with background events from data shown in the left panel and Monte-Carlo signal events in the right.

5.5 Summary

This chapter presents analysis and comparison of the VBF $H \to b\bar{b}$ events attained using trigger-level objects and offline reconstructions in data and Monte-Carlo simulation. The cutflows of the analysis for Monte-Carlo online, Monte-Carlo offline, data online and data offline were studied, and found to show online analysis produced $\sim$ 82% of the results of offline analysis for Monte-Carlo simulations, and $\sim$ 84% for data.

While this is a reduction of event yield for a given analysis sample, the increased trigger rates permitted when applying TLA will result in an increased final event count even given the decrease in efficiency. For current rate increases in TLA analyses [49], this would amount to an increase of $\sim 66\%$ in the final number of events. This estimate of the possible increase does not take into account any limitations on the rate that may result from increased computational demands on either processing or TLA object byte size.

To confirm that the TLA analysis would be possible with the trigger-level objects, the component jets, kinematic properties of the VBF $H \to b\bar{b}$ event and select BDT training variables were investigated. These cases showed consistent behaviour between the online and offline objects in both background data and signal Monte-Carlo simulation, while showing the online rate decrease calculated from the cutflows.

These suggests a full study of VBF $H \to b\bar{b}$ analysis is a feasable proposal. Use of TLA could increase the output rate of the tirggers to statistically significant levels and the objects produced will behave during analysis in a consistent fashion to the offline objects.

Chapter 6

Conclusions

This dissertation contains work done in an attempt to assess the feasibility of using Trigger-Object Level Analysis to improve the statistical significant of searches for the Higgs boson produced via Vector Boson Fusion and decaying to bottom quarks. This feasibility was tested by assessing the performance of trigger-level online analysis to the reconstructed offline analyses at two levels of abstraction for the event, firstly at the resolution of comparing the performance for the individual jet objects that make up a VBF $H \longrightarrow b\bar{b}$ event separate from the VBF $H \longrightarrow b\bar{b}$ topology, and then by performing elements of the full VBF $H \longrightarrow b\bar{b}$ analysis with both online and offline objects to compare the performance.

The analysis was carried out using a vector boson fusion Monte-Carlo simulation sample and 4.63fb^{-1} of data collected by the ATLAS detector during data-taking period D of the 2016 $\sqrt{s}$ = 13TeV Run.

The individual jet objects were shown to be comparable to each other, and could be improved to a closer agreement using standard tools available for ATLAS analyses. The b-jets and non-b-jets that make up a VBF $H \longrightarrow b\bar{b}$event were shown to agree within 1% of each other in their distribution in (η, ϕ) space. The p_{T} distributions of the online and offline jets demonstrated small differences, with offline p_{T} being ~ 5% larger than online p_{T} for the each type of jet object. This p_{T} difference arised from a difference in the jet energy scale calibrations of the online and offline jet objects, and can be rectified in future analyses using standard jet calibration tools.

The b-tagging performance of the individual jet objects was compared, which showed differences in the b-tagging efficiency, c-jet rejection and light-jet rejection between the online and offline objects. The difference in performance was consistent with the expected change in performance coming from applying the 2016 MV2c10 b-tagging algorithm to the offline reconstructed jets, while the b-tagging information for the online jets was calculated using the 2015 MV2c20 algorithm that was operational in the detector at that time. This suggests b-tagging performance of the online objects can be brought into agreement if the b-tagging training variables are preserved on the trigger-level objects, rather than being discarded and leaving only the b-tagging decision.

Comparison was then carried out for the online and offline performance in a VBF $H \rightarrow b\bar{b}$ event phase space. These cuts resulted in a final online event count that was reduced relative to the offline event count, with a final online event fraction of 82% for Monte-Carlo simulation and 84% for data. With the increased trigger rate permitted by using TLA, this would on average increase the final event number by 66% relative to a purely offline analysis.

Finally the VBF $H \rightarrow b\bar{b}$ event specific objects, kinematic quantities and BDT training variables were compared for the online and offline events. For each separate variable, the performance of the online analysis was broadly consistent with respect to the offline analysis, taking into account thre reduction in the number of online events highlighted during the cutflow analysis. These results suggest that TLA analysis in the VBF $H \rightarrow b\bar{b}$ channel will provide increased statistical significance while providing comparable events to the full offline reconstructed analysis.

The work of this dissertation suggests certain additional studies should be carried out prior to approving TLA for the VBF $H \rightarrow b\bar{b}$ channel, that were outside the scope of this analysis. Primarily, the practicalities of applying TLA in the VBF $H \rightarrow b\bar{b}$ channel require assessment. The solutions proposed in this dissertation to improve the agreement between the online and offline objects will increase the size of the trigger-objects output by the detector and may result in additional computational cost in the HLT. These factors may result in a smaller rate increase than assumed based on prior TLA studies and reduce or remove the improvement in rate suggested here.

In addition, the comparitive behaviour of the online and offline objects could have further verification steps. This work did not implement trigger emulation in the Monte-Carlo simulations, and this resulted in discrepancies between the Monte-Carlo and data results for the jet object performance that were explained by the lack of a trigger in the Monte-Carlo simulation. The full VBF $H \rightarrow b\bar{b}$ analysis performed at $\sqrt{s} = 8$TeV carried out a Boosted Decision Tree analysis after the cuts implemented in this dissertation to enhance the VBF $H \rightarrow b\bar{b}$ phase space. Implementing or retraining a BDT was not possible within this dissertation, but would be an informative branch of further work to verify the feasibility. Finally, technical limitations prohibited making use of the full data set produced by the ATLAS detector, as analysis was carried out on a subset only. Greater statistical significance and a more certain statement of similarity could be made by performing the analysis for as large a data collection as possible.

This study on the feasibilty of performing a trigger-object level analysis on the VBF $H \rightarrow b\bar{b}$ channel search for the Higgs boson suggests that the trigger-level objects used for a VBF $H \rightarrow b\bar{b}$analysis are comparable to the offline objects, and that the similarity can be improved with some readily available calibrations and adjustments to the trigger-level objects. Also, the final VBF $H \rightarrow b\bar{b}$ event produced using trigger-level objects will show a worse efficiency compared to offline reconstruction, but with the trigger rate increase afforded by TLA produce more events than an offline analysis, and these events will be comparable in behaviour to the offline reconstruction. There are some additional sections of work relating to implementing and completely verifying the conclusions of this dissertation, but overall trigger-object level analysis is suggested as a feasable analysis strategy in the search for the Higgs boson via the VBF $H \rightarrow b\bar{b}$ channel.

Appendix A

Configuration

This appendix details the files and configuration settings used referenced throughout.

A.1 Files

Table A.1: Full filenames of samples other files used during the analysis

Title	Filename
2016 25ns Good Runs List	`data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects00-02-04_PHYS_StandardGRL_All_Good_25ns.xml`
2016 13TeV `HIGG5D3` sample	`data16_13TeV.{RUN_ID}.physics_Main.merge.DAOD_HIGG5D3.f715_m1620_p2689_tid{TID}`
MC15C `HIGG5D3` derivation Monte-Carlo sample	`mc15_13TeV.341566.PowhegPythia8EvtGen_CT10_AZNLOCTEQ6L1_VBFH125_bb.merge.DAOD_HIGG5D3.e3988_s2726_r7772_r7676_p2719`

A.2 Configurations

Table A.2: Full name configurations used during the analysis

Title	Name
Real Data 20.7 Jet Calibration Recommendations	`JES_data2016_data2015_Recommendation_Dec2016.config`
Monte-Carlo 20.7 Jet Calibration Recommendations	`JES_MC15cRecommendation_May2016.config`
January 2017 MV2c10 *b*-tagging Recommendations	`2016-20_7-13TeV-MC15-CDI-2017-01-31_v1.root`
March 2016 MV2c20 *b*-tagging Recommendations	`2016-Winter-13TeV-MC15-CDI-March10_v1.root`

APPENDIX B

BOOSTED DECISION TREES

This appendix gives a brief description of the definition and use of Boosted Decision Trees (BDT), and provides specific details as to the training of a BDT for a VBF $H \rightarrow b\bar{b}$ analysis.

B.1 Machine Learning

A BDT is a machine learning technique that is applied in analyses to separate signal events from background events. The tree is trained on a particular training sample to nuild the decision logic and then applied to real data as required.

A decision tree as a structure operates by taking variables from the event and creating nodes with child nodes split on ranges of the variables. By assessing the relative signal/background proportions of the child nodes of this split node, the tree can create a split where one side is mostly signal and one mostly background. This process can be applied repeatedly to generate a multiple level tree of decision nodes, iteratively splitting sections of the event dataset. At a final terminating leaf node of the tree, the proportions of the signal and background events in the node will label it as a signal node or a background node.

This structure once trained, can be used to label a measured event by moving down the tree and evaluating each decision before a leaf node is reached in order to categorise the event. The boosting of a decision tree refers to the process of applying weights to the events. The tree will be iteratively produced, reweighting any misclassified events at each iterative stage to produce a more refined final tree [58]. Such structures are used throughout modern physics analyses at ATLAS [59].

B.2 VBF $H \to b\bar{b}$ BDT Training

A detailed description of the BDT training that should be carried out for a VBF $H \to b\bar{b}$ search is given in Ref. [60]. Here we summarise the event variables used for training the BDT on the VBF $H \to b\bar{b}$ events.

Table B.1: BDT Variables used in training for the VBF $H \to b\bar{b}$ analysis.

Variable	Description
M_{jj}	Invariant mass of the VBF jet pair.
$p_{\mathrm{T}jj}$	Transverse momentum of the VBF jet pair
$\cos\theta$	Cosine of the polar angle of the cross product of the VBF jet momenta in the Higgs rest frame.
$Max(\eta)$	$max(\lvert\eta_{j1}\rvert, \lvert\eta_{j2}\rvert)$ Maximum of the two absolute pseudorapidity values for the VBf jets.
$\eta*$	$\frac{1}{2}(\lvert\eta_{j1}\rvert + \lvert\eta_{j2}\rvert - \lvert\eta_{b1}\rvert - \lvert\eta_{b2}\rvert)$ Average pseudorapidity difference between the VBF and signal jets.
$min\mathit{\Delta}R_{j1}$	Minimum (η, ϕ) separation between the leading VBF jet and the closest other jet.
$min\mathit{\Delta}R_{j2}$	Minimum (η, ϕ) separation between the sub-leading VBF jet and the closest other jet.
QuarkGluonTagger(j_1)	Number of tracks associated with the leading VBF jet [61].
QuarkGluonTagger(j_2)	Number of tracks associated with the sub-leading VBF jet.
p_{T} Balance	Ratio of vectorial and scalar sum of signal and VBF jets: $\frac{\boldsymbol{p}_{\mathrm{T}\boldsymbol{j1}}+\boldsymbol{p}_{\mathrm{T}\boldsymbol{j2}}+\boldsymbol{p}_{\mathrm{T}\boldsymbol{b1}}+\boldsymbol{p}_{\mathrm{T}\boldsymbol{b2}}}{p_{\mathrm{T}j1}+p_{\mathrm{T}j2}+p_{\mathrm{T}b1}+p_{\mathrm{T}b2}}$.
$\mathit{\Delta}M_{jj}$	Difference in the largest invariant mass from all jet pairs and the invariant mass of the VBF jet pair

Bibliography

[1] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593 [hep-ph].

[2] LHC Higgs Cross Section Working Group Collaboration, J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, arXiv:1307.1347 [hep-ph].

[3] J. Pequenao, "Computer generated image of the whole ATLAS detector." https://cds.cern.ch/record/1095924. Accessed 03/09/2017.

[4] ATLAS Collaboration, M. Aaboud et al., *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. **C77** no. 5, (2017) 317, arXiv:1611.09661 [hep-ex].

[5] Particle Data Group Collaboration, C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. **C40** no. 10, (2016) 100001.

[6] ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[7] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588.

[8] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[9] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367–377.

[10] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531–533. [,648(1963)].

[11] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[12] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[13] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[14] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132–133.

[15] ATLAS Collaboration, G. Aad et al., *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) 1–29, `arXiv:1207.7214 [hep-ex]`.

[16] CMS Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30–61, `arXiv:1207.7235 [hep-ex]`.

[17] NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, `arXiv:1410.8849 [hep-ph]`.

[18] J. C. Collins, *Light cone variables, rapidity and all that*, `arXiv:hep-ph/9705393 [hep-ph]`.

[19] M. H. Seymour and M. Marx, *Monte Carlo Event Generators*, pp. , 287–319. 2013. `arXiv:1304.6677 [hep-ph]`. `https://inspirehep.net/record/1229804/files/arXiv:1304.6677.pdf`.

[20] B. Andersson, G. Gustafson, G. Ingelman, and T. SjÃűstrand, *Parton fragmentation and string dynamics*, Physics Reports **97** no. 2, (1983) 31 – 145. `http://www.sciencedirect.com/science/article/pii/0370157383900807`.

[21] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. **B238** (1984) 492–528.

[22] T. SjÃűstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159–177, `arXiv:1410.3012 [hep-ph]`.

[23] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, `arXiv:0811.4622 [hep-ph]`.

[24] C. Oleari, *The POWHEG-BOX*, Nucl. Phys. Proc. Suppl. **205-206** (2010) 36–41, `arXiv:1007.3893 [hep-ph]`.

[25] S. Asai et al., *Prospects for the search for a standard model Higgs boson in ATLAS using vector boson fusion*, Eur. Phys. J. **C32S2** (2004) 19–54, `arXiv:hep-ph/0402254 [hep-ph]`.

[26] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[27] *LEP design report*. CERN, Geneva, 1983. https://cds.cern.ch/record/98881. By the LEP Injector Study Group.

[28] *LEP design report*. CERN, Geneva, 1984. https://cds.cern.ch/record/102083. Copies shelved as reports in LEP, PS and SPS libraries.

[29] Y. Koshiba et al., *Luminosity Increase in Laser-Compton Scattering by Crab Crossing Method*, in *Proc. of International Particle Accelerator Conference (IPAC'17), Copenhagen, Denmark, 14ÃćÂĂÂŞ19 May, 2017*, pp. , 902–904. JACoW, Geneva, Switzerland, May, 2017. http://jacow.org/ipac2017/papers/mopva023.pdf. https://doi.org/10.18429/JACoW-IPAC2017-MOPVA023.

[30] ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol. 1*,.

[31] H. Wilkens and the ATLAS LArg Collaboration, *The ATLAS Liquid Argon calorimeter: An overview*, Journal of Physics: Conference Series **160** no. 1, (2009) 012043. http://stacks.iop.org/1742-6596/160/i=1/a=012043.

[32] for the ATLAS collaboration Collaboration, A. M. Henriques Correia, *The ATLAS Tile Calorimeter*, Tech. Rep. ATL-TILECAL-PROC-2015-002, CERN, Geneva, Mar, 2015. https://cds.cern.ch/record/2004868.

[33] A. Artamonov, D. Bailey, G. Belanger, M. Cadabeschi, T. Y. Chen, V. Epshteyn, P. Gorbounov, K. K. Joo, M. Khakzad, V. Khovanskiy, P. Krieger, P. Loch, J. Mayer, E. Neuheimer, F. G. Oakham, M. O'Neill, R. S. Orr, M. Qi, J. Rutherfoord, A. Savine, M. Schram, P. Shatalov, L. Shaver, M. Shupe, G. Stairs, V. Strickland, D. Tompkins, I. Tsukerman, and K. Vincent, *The ATLAS Forward Calorimeter*, Journal of Instrumentation **3** no. 02, (2008) P02010. http://stacks.iop.org/1748-0221/3/i=02/a=P02010.

[34] ATLAS Collaboration Collaboration, M. zur Nedden, *The Run-2 ATLAS Trigger System: Design, Performance and Plan*, Tech. Rep. ATL-DAQ-PROC-2016-039, CERN, Geneva, Dec, 2016. https://cds.cern.ch/record/2238679.

[35] R. Achenbach et al., *The ATLAS level-1 calorimeter trigger*, JINST **3** (2008) P03001.

[36] ATLAS Collaboration Collaboration, *Trigger Menu in 2016*, Tech. Rep. ATL-DAQ-PUB-2017-001, CERN, Geneva, Jan, 2017. https://cds.cern.ch/record/2242069.

[37] G. P. Salam, *Towards Jetography*, Eur. Phys. J. **C67** (2010) 637–686, arXiv:0906.1833 [hep-ph].

[38] R. Atkin, *Review of jet reconstruction algorithms*, J. Phys. Conf. Ser. **645** no. 1, (2015) 012008.

[39] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[40] O. Lundberg, *Calibration Systems of the ATLAS Tile Calorimeter*, pp. , 399–402. 2012. arXiv:1212.3676 [physics.ins-det]. https://inspirehep.net/record/1207575/files/arXiv:1212.3676.pdf.

[41] G. Pospelov and the Atlas Hadronic Calibration Group, *The overview of the ATLAS local hadronic calibration*, Journal of Physics: Conference Series **160** no. 1, (2009) 012079. http://stacks.iop.org/1742-6596/160/i=1/a=012079.

[42] Z. Marshall and the Atlas Collaboration, *Simulation of Pile-up in the ATLAS Experiment*, Journal of Physics: Conference Series **513** no. 2, (2014) 022024. http://stacks.iop.org/1742-6596/513/i=2/a=022024.

[43] *Tagging and suppression of pileup jets with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2014-018, CERN, Geneva, May, 2014. https://cds.cern.ch/record/1700870.

[44] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11** no. 04, (2016) P04008, arXiv:1512.01094 [hep-ex].

[45] *Expected performance of the ATLAS b-tagging algorithms in Run-2*, Tech. Rep. ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul, 2015. https://cds.cern.ch/record/2037697.

[46] ATLAS Collaboration, *Optimisation of the ATLAS b-taggingperformance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012 (2016). https://cds.cern.ch/record/2160731.

[47] ATLAS Collaboration Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, Tech. Rep. ATLAS-CONF-2011-102, CERN, Geneva, Jul, 2011. http://cds.cern.ch/record/1369219.

[48] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444–450.

[49] ATLAS Collaboration, T. A. collaboration, *Search for light dijet resonances with the ATLAS detector using a Trigger-Level Analysis in LHC pp collisions at* $\sqrt{s} = 13$ *TeV*,.

[50] I. Antcheva, M. Ballintijn, B. Bellenot, M. Biskup, R. Brun, N. Buncic, P. Canal, D. Casadei, O. Couet, V. Fine, L. Franco, G. Ganis, A. Gheata, D. G. Maline, M. Goto, J. Iwaszkiewicz, A. Kreshuk, D. M. Segura, R. Maunder, L. Moneta, A. Naumann, E. Offermann, V. Onuchin, S. Panacek, F. Rademakers, P. Russo, and M. Tadel, *ROOT âĂŤ A C++ framework for petabyte data storage, statistical analysis and visualization*, Computer Physics Communications **180** no. 12, (2009) 2499 – 2512. http://www.sciencedirect.com/science/article/pii/S0010465509002550. 40 YEARS OF CPC: A celebratory issue focused on quality software for high performance, grid and novel computing architectures.

[51] J. Catmore, J. Cranshaw, T. Gillam, E. Gramstad, P. Laycock, N. Ozturk, and G. A. Stewart, *A new petabyte-scale data derivation framework for ATLAS*, Journal of Physics: Conference Series **664** no. 7, (2015) 072007. http://stacks.iop.org/1742-6596/664/i=7/a=072007.

[52] R. Seuster, M. Elsing, G. A. Stewart, and V. Tsulaia, *Status and Future Evolution of the ATLAS Offline Software*, Journal of Physics: Conference Series **664** no. 7, (2015) 072044. http://stacks.iop.org/1742-6596/664/i=7/a=072044.

[53] "Worldwide lhc computing grid website." Http://wlcg.web.cern.ch/. Accessed: 2018-01-09.

[54] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, arXiv:hep-ph/0201195 [hep-ph].

[55] ATLAS Collaboration, G. Aad et al., *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv:1406.3660 [hep-ex].

[56] ATLAS Collaboration, M. Aaboud et al., *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. **D96** no. 7, (2017) 072002, arXiv:1703.09665 [hep-ex].

[57] A. Schwartzman, *Jet energy calibration at the LHC*, Int. J. Mod. Phys. **A30** no. 31, (2015) 1546002, arXiv:1509.05459 [hep-ex].

[58] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, *Boosted decision trees, an alternative to artificial neural networks*, Nucl. Instrum. Meth. **A543** no. 2-3, (2005) 577–584, arXiv:physics/0408124 [physics].

[59] M. Paganini, *Machine Learning Algorithms for b-Jet Tagging at the ATLAS Experiment*, in *18th International Workshop on Advanced Computing and Analysis*

Techniques in Physics Research (ACAT 2017) Seattle, WA, USA, August 21-25, 2017. 2017. arXiv:1711.08811 [hep-ex]. https://inspirehep.net/record/1638366/files/arXiv:1711.08811.pdf.

[60] ATLAS Collaboration, M. Aaboud et al., *Search for the Standard Model Higgs boson produced by vector-boson fusion and decaying to bottom quarks in $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, JHEP **11** (2016) 112, arXiv:1606.02181 [hep-ex].

[61] J. Gallicchio and M. D. Schwartz, *Quark and Gluon Tagging at the LHC*, Phys. Rev. Lett. **107** (2011) 172001, arXiv:1106.3076 [hep-ph].