Trigger Level Analysis

Andrew James Strange

School of Physics and Astronomy

2017 September

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

Contents

Declaration

This is the declaration. This is not too long, honest!

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 $SU(3)_c \times SU(2)_L \times U(1)_\gamma$ 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 1.3). 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 1.1 and 1.2, 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)). This

Most matter consists of the observable first generation of the up and down quarks and

the electron, 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 neutral leptons experiencing only the weak interaction.

Table 1.1: Spin-$\frac{1}{2}$ fermions: quarks q [1]

Generation	Flavour	Charge / e	Mass / GeV
1	Up u	+2/3	0.002
	Down d	-1/3	0.005
2	Charm c	+2/3	1.28
	Strange s	-1/3	0.096
3	Top t	+2/3	173.1
	Bottom b	-1/3	4.18

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

Generation	Flavour	Charge / e	Mass / MeV
1	Electron e	-1	0.511
	Electron Neutrino ν_e	0	~0
2	Muon μ	-1	105.658
	Muon Neutrino ν_μ	0	~0
3	Tau τ	-1	1776.86
	Tau Neutrino ν_τ	0	~0

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* ($u\bar{d}$). When a high energy hadron is produced, the interaction of the strong force on the quarks results is a collimated *jet* of hadrons that freeze out of the initial hadron. This process is described in more detail in Section 1.2.2.3.

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 1.3. In addition to the forces, particles acquire mass by coupling to the *Higgs* field via the spin-0 Higgs boson [2–4], which is covered in more detail in Section 1.1.3.

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. 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. [1]

Interaction	Particle	Charge / e	Mass / GeV	Strength (α)
Strong	Gluon g	0	0	~ 1
Weak (Charged Current)	W^+	1	80.4	
	W^-	-1	80.4	10^{-6}
Weak (Neutral Current)	Z	0	91.2	
Electromagnetic (EM)	Photon γ	0	0	$\frac{1}{137}$

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 while quarks are the only fermion to feel the strong interaction, gluons can self-couple.

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 [5–7], which corresponds to the $SU(2)_L \times U(1)_\gamma$ 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 $SU(2)_L \times U(1)_\gamma$ group produces four gauge bosons which mix to produce the more recognisable γ, W^+, W^- and Z bosons. The unified force couples to weak isospin, which allows self-coupling between the massive vector bosons, but not the photon as it does not carry electric charge.

While the weak interaction acts on both quarks and leptons, the quark sector is affected by

the distinction between the mass and flavour eigenstates of quarks. The physically observed flavour eigenstates are distinct from the quark eigenstates of the weak interactions, which are superpositions of the mass eigenstates. The effect of this quark mixing in the weak interaction is that different flavour changing interactions have different strengths. The mixing of the mass eigenstates (q) into weak eigenstates (q') is described by the Cabbibo-Kobayashi-Makasawa matrix [8, 9]:

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

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, which follows from the Klein-Gordon equation:

$$\frac{\partial^2 \psi}{\partial t^2} = (\nabla^2 - m^2)\psi \tag{1.2}$$

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 introducing a scalar field that interacts 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 addition to the Lagrangian is still required to preserve the symmetry of the system and respect the gauge invariance, but is also required to have a non-zero expectation value for the field in the vacuum or ground state of space. The Higgs mechanism introduces the scalar field ϕ which has a potential energy $V(\phi)$:

$$V(\phi) = a\phi^4 - b\phi^2 \tag{1.3}$$

This results in an equilibrium point ($\phi = 0$) that respects the symmetry, but is inherently unstable, with an infinite set of degenerate non-zero minima at $|\phi^2| = \frac{b}{2a}$ 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*. Confirmation of the Higgs boson as part of the SM was only achieved relatively recently [10, 11], where a spin-0 boson consistent with the SM Higgs was observed. Subsequent measurements made have provided agreement on the new particle as the Higgs boson with a mass of 125.09 GeV [1]. Section 1.3 covers in more detail the production and behaviour of the Higgs boson in collider experiments.

1.2 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 detectors. 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 [12].

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 $pp \to X$, all possible combinations of incoming partons must be summed over and the momentum fractions integrated over while accounting for the PDFs:

$$\sigma = \sum_{i,j=q,g} \int dx_1 dx_2 f_i(x_1, Q^2) f_j(x_2, Q^2) \hat{\sigma}(Q^2) \tag{1.4}$$

1.2.1 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 [13].

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:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{1.5}$$

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:

$$\eta = -\ln \tan \frac{\theta}{2} \tag{1.6}$$

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

1.2.2 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 [14].

1.2.2.1 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.

1.2.2.2 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.

1.2.2.3 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 [15] or a Cluster model [16]. These steps are models, and not calculations as to how the partons combine, as to calculations being prohibited by the breakdown of perturbation theory.

1.2.2.4 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.

1.2.2.5 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.

1.2.3 Monte-Carlo Software

There is a broad selection of software tools for evaluating pp collisions, from general purpose simulations like PYTHIA [17] or SHERPA [18] which are used to evaluate the complete process, to more specific tools like POWHEG [19] 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.

1.3 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.

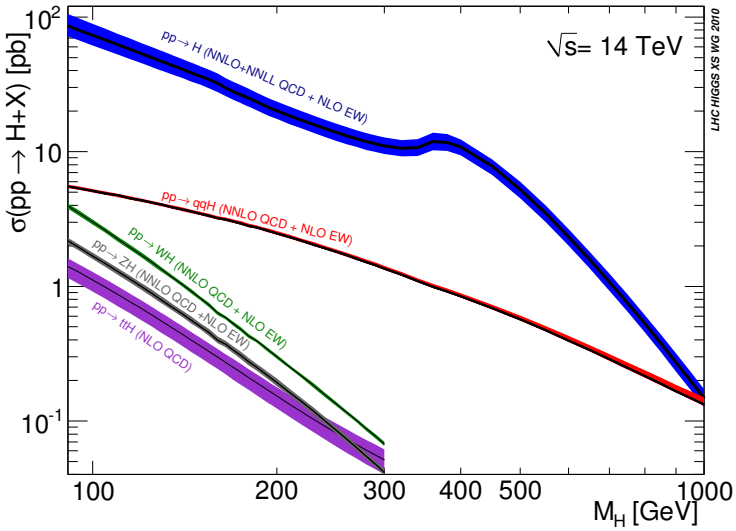


Figure 1.1: SM Higgs Production cross section for $\sqrt{s} = 14$ TeV. $pp \rightarrow H$ corresponds to gluon-gluon fusion production and $pp \rightarrow qqH$ vector boson fusion. [20]

1.3.1 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 1.1, with the

second largest cross-section arising from Vector Boson Fusion (VBF, Section 1.3.2). 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) [20]. The lowest order Feynmann diagrams for these processes are shown in Figure 1.2.

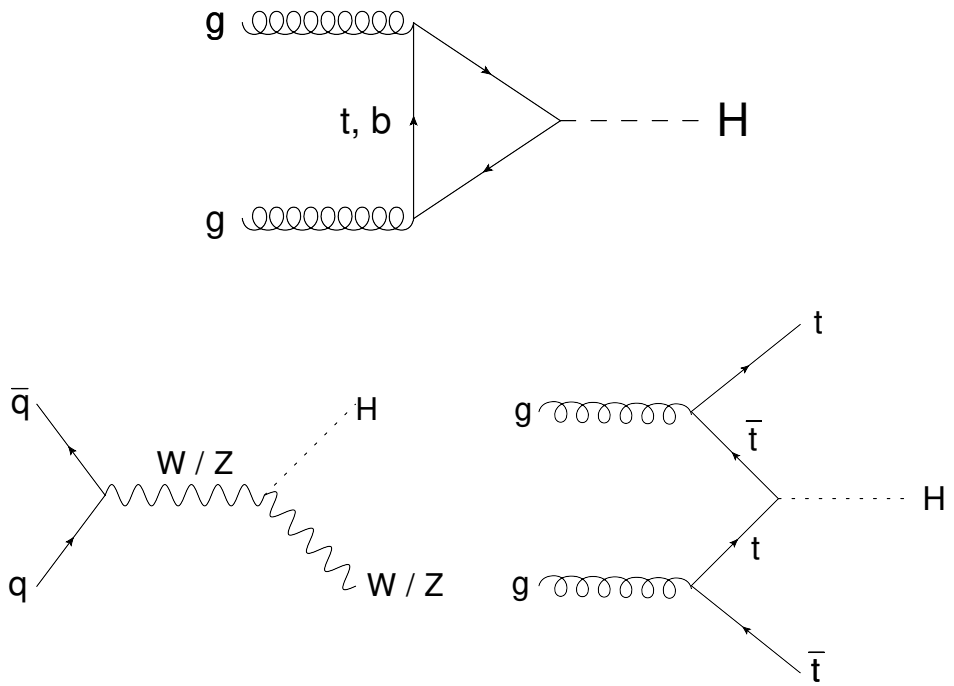


Figure 1.2: Lowest order Feynmann diagrams for gluon-gluon fusion ($gg \rightarrow H$), W/Z associated production (WH/ZH) and top anti-top associated production ($t\bar{t}H$).

1.3.1.1 Gluon-gluon Fusion

The dominant production mechanism for the Higgs boson in hadron colliders is the $gg \rightarrow H$ 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 [20].

1.3.2 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 1.3. 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 [21]. 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.

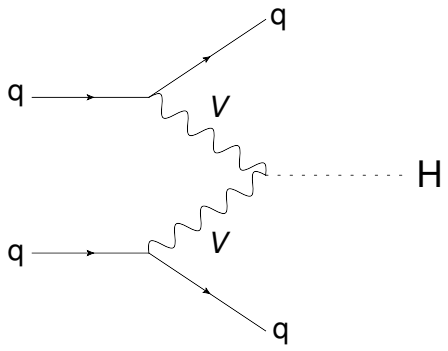


Figure 1.3: Feynmann diagram for the production of a Higgs boson via Vector Boson Fusion, where q denotes any quark or antiquark

1.3.3 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 1.4.

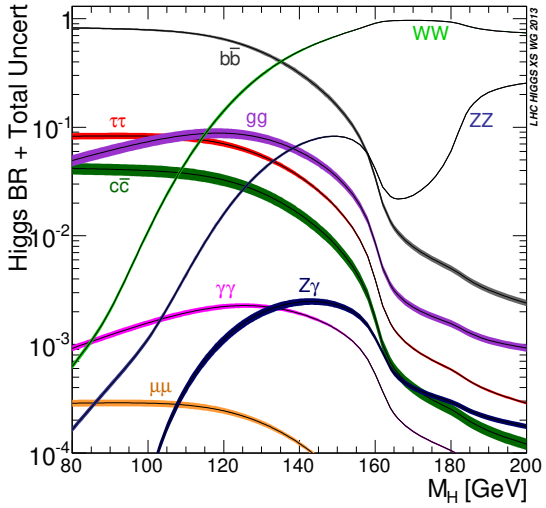


Figure 1.4: Higgs decay branching ratios for the low mass region with their uncertainties [22].

While observations consistent with the Standard Model Higgs boson have been made for the $H \to \gamma\gamma$, $H \to ZZ$, $H \to W^+W^-$ and $H \to \tau^+\tau^-$ channels, observation of th $H \to bb$ 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.

1.3.4 VBF Searches

Searches fo the VBF $H \to b\bar{b}$ 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 $H \to b\bar{b}$ events are indistinguishable from the $gg \to H$ 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 heavy ions or protons at a *centre-of-mass* energy $\sqrt{s} = 14\text{TeV}$ and a peak *luminosity* of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ [23]. 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 [24, 25]. 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 $\int L dt$ 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 $\sqrt{s}$ = 13TeV. Over the course of the 2016 data-taking the LHC provided an integrated luminosity of $\sim$ 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 [26].

2.2 The ATLAS Detector

The ATLAS detector [27] 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 2.1. 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 2.1.

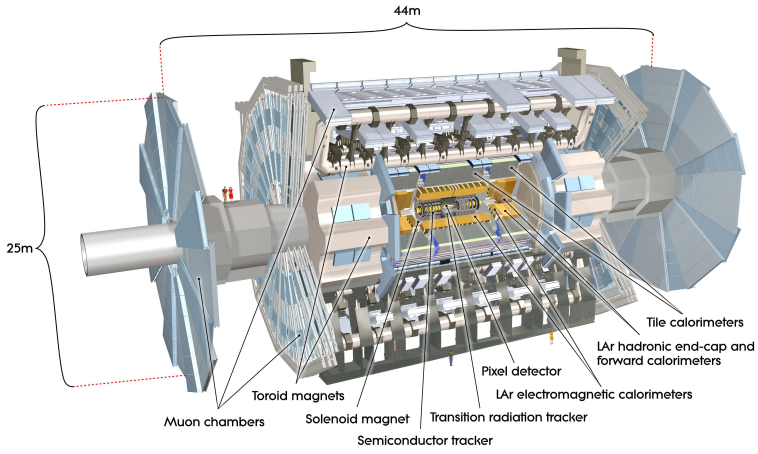


Figure 2.1: Schematic cut-away of the ATLAS detector [28].

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 [29] (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 2.1), 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 [27].

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 [27].

The EM calorimeter [30] 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 3 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 [**?**] 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 [31] is contained within the end-cap cryostat (The FCal is omitted from Figure 2.1) 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 $O(10)$ns, while the precision momentum measurement is carried out in Monitored Drift Tubes arranged in layers [27].

Table 2.1: Performance goals and operational ranges for the principal components of the ATLAS detector. [27]

System	Component	η Coverage	Resolution
Tracking		$0 < \|\eta\| < 2.5$	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$
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		$0 < \|\eta\| < 2.7$	$\sigma_{p_T}/p_T = 10\%$ *at* $p_T = 1$ TeV

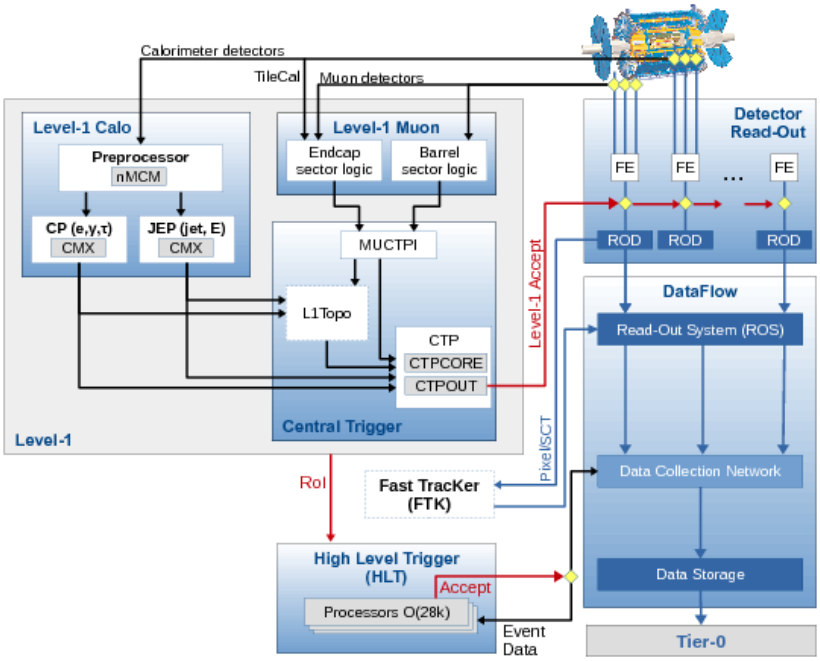


Figure 2.2: Schematic plot of the ATLAS Trigger and Data acquisition system [33].

2.3 Trigger and data acquisition

When operating at the design luminosity, the LHC produces a bunch-crossing rate of 40 MHz [32]. 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 2.2. 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 [34] 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 [32]. 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 [**?**], 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 event storage handled by the trigger chains and prescaling, only select sections if the overall output dataset 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 3, 4 and 5.

2.5 Object Reconstruction

2.5.1 Jets

As discussed in Section 1.2.2.3 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 [35]. Once a cone is defined, any constituents contained within the cone are removed from consideration and the process repeats. The alternate 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

$$d_{ij} = min(p^a_{\mathrm{T}i}, p^a_{\mathrm{T}j})\frac{\Delta R^2_{ij}}{R} \tag{2.2}$$

where a is a particular exponent for a given algorithm, R is the radius parameter of the final reconstructed jet size and ΔR_{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 [36] and is given by

$$d_{iB} = p^a_{\mathrm{T}i} \tag{2.3}$$

The principal algorithm used for jet reconstruction at ATLAS is the anti-k_t algorithm [37], 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 [38] calorimeter cells according to the Electromagnetic (EM) scale, or by applying Local Cell (LC) corrections [39] 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 1.2.2.5 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 [40]. 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 [41].

2.5.2 *b*-Tagging

Hadrons containing a b-quark tend to feature a signature topology as a result of the long lifetime of . 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 [42].

Identification of jets containing in ATLAS is based on combining the output of three separate lifetime-based b-tagging algorithms [43]: Impact Parameter based algorithms (IP2D and IP3D, Section 2.5.2.1), Secondary Vertex based (SV, Section 2.5.2.2) and Decay Chain based (JetFitter, Section 2.5.2.3) into a multivariate discriminant (MV2, Section 2.5.3) 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 2.2.1). 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 [41, 44].

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, ϕ) 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, ϕ) and the primary vertex in the longitudinal plane [45]. 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 if 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. [42, 44]

2.5.2.2 SV1: Secondary Vertex Finding algorithm

The secondary vertex algorithm uses the decay products of the b-hadronto reconstruct a distinct secondary vertex [45]. 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 [42, 44].

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 [46]. 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 [42].

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-jetsfrom 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-jetcomposition of the training sample; for MV2c20 the b-jetsare designated as signal jets where a mixture of 80% light jets and 20% c-jetswas designated as background [43].

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-jetselection 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-jetefficiency (corresponding to *looser* b-jetselection) 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-jetslimiting physics analyses more the light-jets. This change to enhance the c-jetrejection meant that for the MV2c10, the c-jetfraction 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*-jetrejection, so was chosen as the standard *b*-taggingalgorithm for 2016 analyses [44].

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 $\times$ 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) [47].

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 [47] a partial TLA event was 5% of the size of a full detector readout.

Chapter 3

Event Selection

This chapter is describes the selection criteria for real and simulated event data, 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 analysis towards the typical VBF $H \rightarrow b\bar{b}$ final state described in Section 1.3.2.

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 [48], 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 [49].

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 [50]. The computation framework used for analysis of the xAOD dat is the internally

developed Analysis Base suite of tools. This analysis uses Analysis Base 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 [51] to provide the necessary hardware capacity.

3.2 Event weights

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:

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

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.

3.3 Samples

Real event data was taken from the 2016 13 TeV run, with Data Period D used owing to limited storage space on analysis computing facilities. The `HIGG5D3` derivation was used for the data and Monte-Carlo samples, with a full list of tags given in Appendix A. This analysis used the all year 25ns Good Runs List (Table A.1), resulting a a data luminosity of 4.6312fb^{-1}. The simulated VBF sample (Table A.1) was produced during the MC15c production period. This sample was produced using the NLO generator POWHEG configured using the CTEQ6L1 [52] set of PDFs and interfaced with PYTHIA8 tuned to AZNLO [53].

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 2.5.1. 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, given the high p_{T} cuts required for an event, as discussed in Section 3.5, all jets were required to have $p_{\mathrm{T}} > 45$GeV.

In selecting the trigger level offline jets, firstly all *split*-jets that pass the trigger are retrieved from the trigger chain. Any duplicates, determined through ΔR matching, are removed and the b-tagging information, which is stored in a separate xAOD container, is associated with the jets. Following this all L1 trigger jets are retrieved, which do not possess b-tagging information. The full set of L1 jets is compared to the *split*-jets and any duplicates are removed from the L1 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.

When searching for b-jets or forward jets in the complete jet collection for online jets, only the *split*-jets can be considered for they are associated with b-taggininformation. Both *split* and *nonsplit*-jets can be considered for the VBF jet during jet assignment as describing in Section 3.5.

3.4.1 b-jets

The details of b-jetidentification are covered in Section 2.5.2. 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.

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

Following from the description of the VBF $H \to b\bar{b}$ events in Section 1.3.2, target events are selected by requiring two central b-jets which form the Higgs candidate and two high p_{T} VBF jets. 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 one forward jet. In this study, the online trigger level jets could not be extracted for the specific trigger chains used previously for the *four-central* channel, so analysis focuses on the *two-central* channel.

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 $E_{\mathrm{T}} > 40$GeV and $|\eta| < 2.5$. In addition, a second central jet ROI with $E_{\mathrm{T}} > 25$ and a forward jet ROI with $E_{\mathrm{T}} > 20$GeV and $3.1 < |\eta| < 4.9$ are both required. At the HLT level, one central jet b-tagged at the *tight* working point with $p_{\mathrm{T}} > 80$GeV, and a jet with $p_{\mathrm{T}} > 60$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 $p_{\mathrm{T}} > 95$GeV which was b-tagged at the *tight* working point and one additional jet with $p_{\mathrm{T}} > 70$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 bb pair was required to exceed 160 GeV. This cut is to remove kinetic 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 $p_{\mathrm{T}bb}$, with the $p_{\mathrm{T}bb}$ distribution featuring a peak as a result of the individual jet p_{T} requirements. 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 $H \to b\bar{b}$ 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 and designated b-jetsb_1 and b_2 with respect to individual p_{T} . 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 and labeled j_1, j_2 according to p_{T} .

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

In a full analysis, the signal is extracted from the results using a Boosted Decision Tree trained to extract the VBF $H \to b\bar{b}$ events from the $gg \to H$ contributions. 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 $H \rightarrow b\bar{b}$ 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 1.2.1) 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 4.1.

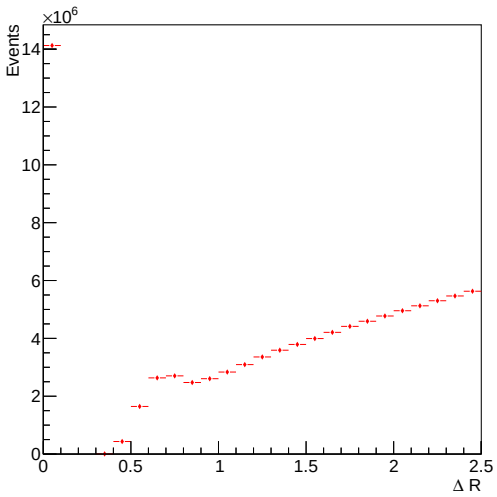


Figure 4.1: Plot of ΔR values for all online/offline jet pairs taken from the Monte-Carlo data. The large spike at ~ 0 accounts for matching jets, with the higher ΔR Values corresponding to differing jet pairs.

To compare the online and offline jets, the ratio of the difference in value for (significant) jet properties between the matched jets were evaluated. These values were calculated for jet feature X using:

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

where $X_{Offline}$ is the value of the feature on the offline jet, and X_{Online} is from the HLT jet. Specific categories of jets within the event were compared to ensure the most relevant jets for the VBF $H \rightarrow b\bar{b}$ analysis were comparable.

The performance of the online and offline jet features was tested bearing in mind the final state products of the VBF $H \rightarrow b\bar{b}$ interaction, while considering the rapidity of the produced jets. The jets used to produce these plots were taken from all analysed Monte-Carlo events and all real data events where the trigger was passed, but the additional VBF $H \rightarrow b\bar{b}$ requirements mentioned in Section 3.5 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-taggingworking 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.

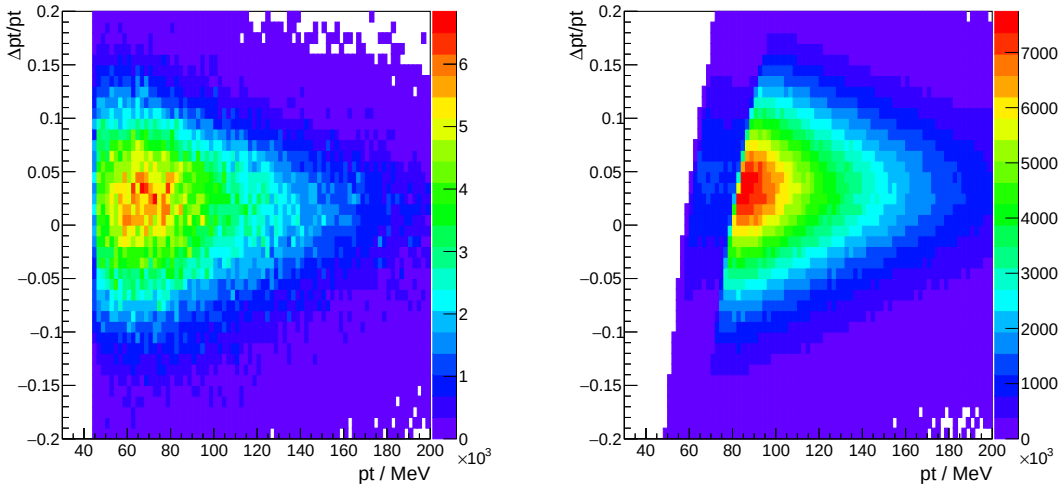


Figure 4.2: $\Delta p_{\mathrm{T}\ ratio}$ for the leading p_{T} b-jet against p_{T} of the offline b-jet, plotted for Monte-Carlo simulated results (left) and real data (right).

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 $\frac{\Delta p_T}{p_T} \sim 0$ and the two plots show a comparable drop off in p_T distribution. The distinctive curve present in the real data is the result of the trigger (Section 3.5 being applied to each event, which was not applied in the Monte-Carlo data. The real data was required to contain at least one jet with a $p_T > 80$GeV which results in the steep drop-off below this cut value.

The distribution of the $\frac{\Delta p_T}{p_T}$ about 0 can be shown by taking a slice across the distribution for a representative p_T value. The $\frac{\Delta p_T}{p_T}$ values were also split into η bands to study performance at different points in the detector. For the leading b-jet, this is constrained to be within the region of the detector where is available.

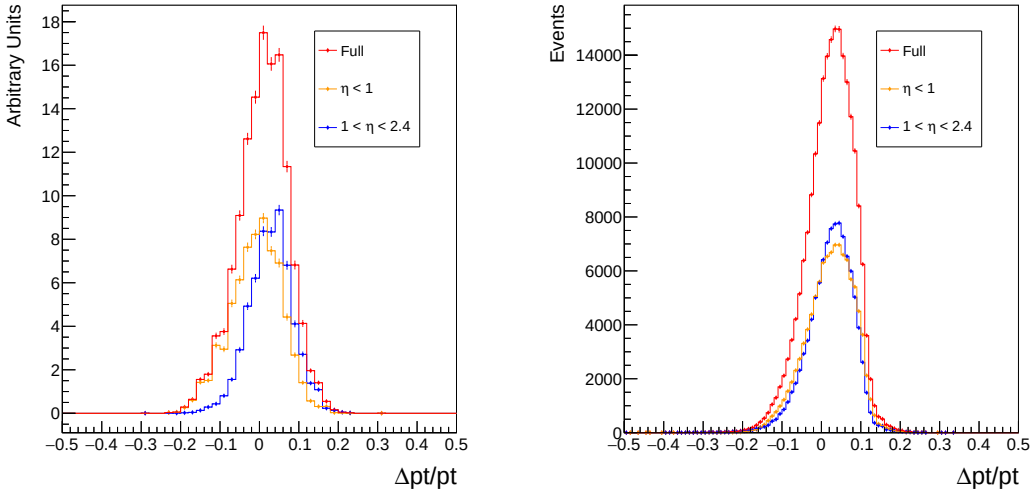


Figure 4.3: $\frac{\Delta p_T}{p_T}$ distribution for the leading b-jetwith 89 $< p_T <$ 91 GeV. The distributions for all events and events split by η region are shown. Monte-Carlo results are shown on the left and real data on the right.

The results show similar profiles between the Monte-Carlo and Data events for $\frac{\Delta p_T}{p_T}$. Both plots show the offline p_T values to be consistently 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 showed a slight difference in $\frac{\Delta p_T}{p_T}$ value as the central η range peaked at ~ 0. The breadth of these distributions is quite large, with both Data and Monte-Carlo showing $\pm 10\%$ in $\frac{\Delta p_T}{p_T}$.

These comparisons can be carried out for other jet properties (η, ϕ) to confirm the offline and online jets are behaving in a similar fashion.

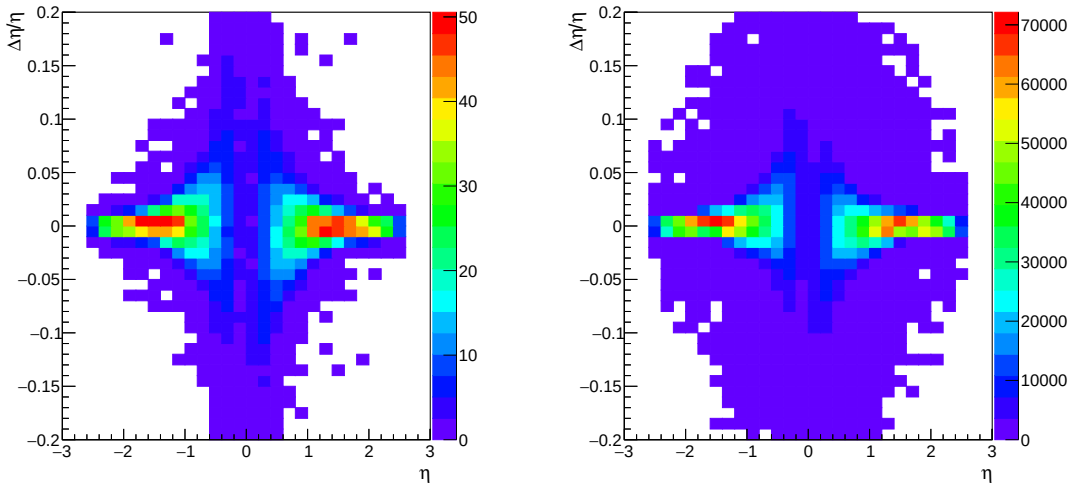


Figure 4.4: $\frac{\Delta\eta}{\eta}$ for the leading b-jet, for Monte-Carlo events (left) and real data (right).

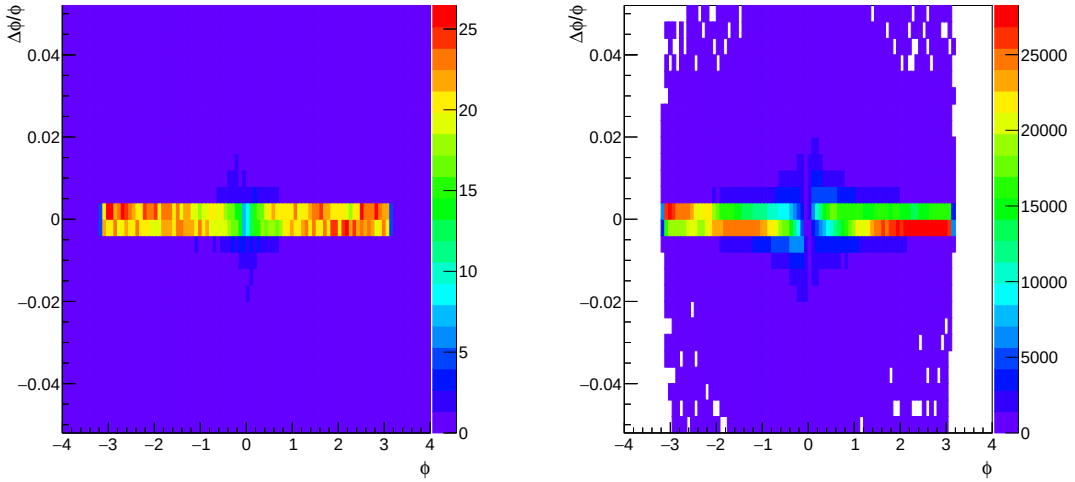


Figure 4.5: $\frac{\Delta\phi}{\phi}$ for the leading b-jet, for Monte-Carlo events (left) and real data (right).

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. In both cases the median $\frac{\Delta X}{X}$ value is ~ 0 and the breadth of the distribution is less than 1% of the value.

4.2 Leading Non *b*-jets

For VBF $H \rightarrow b\bar{b}$ the pair of high p_{T} forward jets is the other significant feature, so the offline/online performance in the leading non-*b*-jetwas studied.

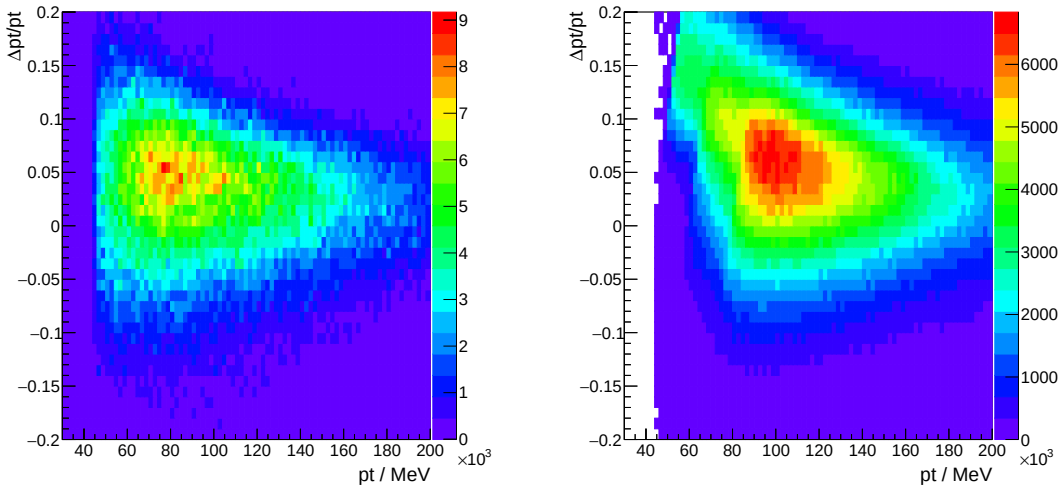


Figure 4.6: $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ for the leading p_{T} non *b*-jet against p_{T} of the offline jet, plotted for Monte-Carlo simulated results (left) and real data (right).

The leading non *b*-jet shows a similar situation to the leading *b*-jet. At the peak of the $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$ distribution the p_{T} of the offline jet is within 5% of the online jet, and the overall shape of the distribution is comparable between the Monte-Carlo and real Data events. The results could also be split across η regions, with the added ability to study the forward regions of the detector.

Both Monte-Carlo and real data show offline jets to be consistently higher p_{T} than online 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 eta 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, showing that the online p_{T} exceeded the offline p_{T} . In addition, the relative proportions of the three η bands differ. The central region of the Monte-Carlo distribution has a flattened peak and peaks at a negative value, showing the offline p_{T} was less than the online, compared toa positive value for the real data. The Both the MC and the Data show that the offset in p_{T} value is worse for the jets in the forward region than in the central regions of the detector, with a significantly higher median $\frac{\Delta p_{\mathrm{T}}}{p_{\mathrm{T}}}$.

The other topological variables can also be compared.

As with the *b*-jets these other variables offline and online jets produce nearly identical results, with the distribution of $\frac{\Delta X}{X}$ firmly centred around 0 and a breadth of less than 1%.

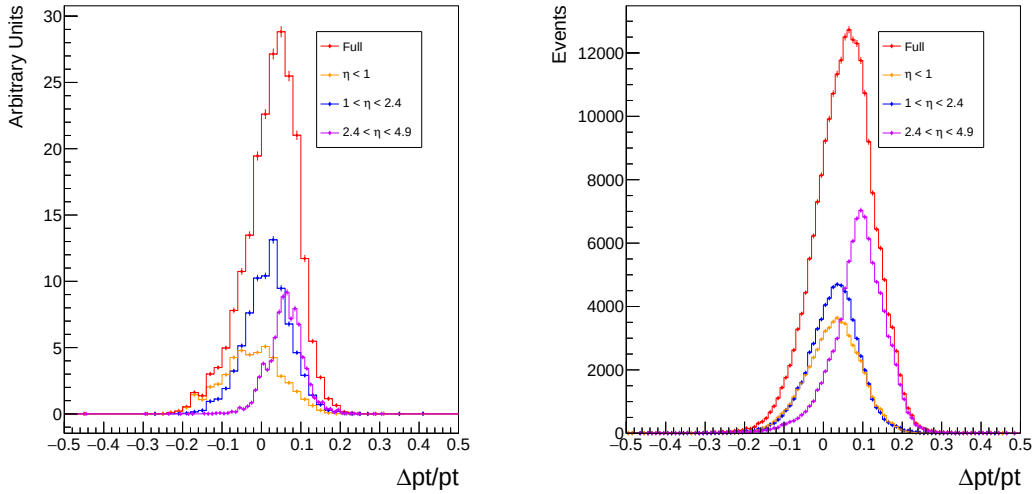


Figure 4.7: $\frac{\Delta p_T}{p_T}$ distribution for the leading non b-jet with 89 $< p_T <$ 91 GeV plotted for Monte-Carlo simulated results (left) and real data (right). The distributions for all events and events split by η region are shown.

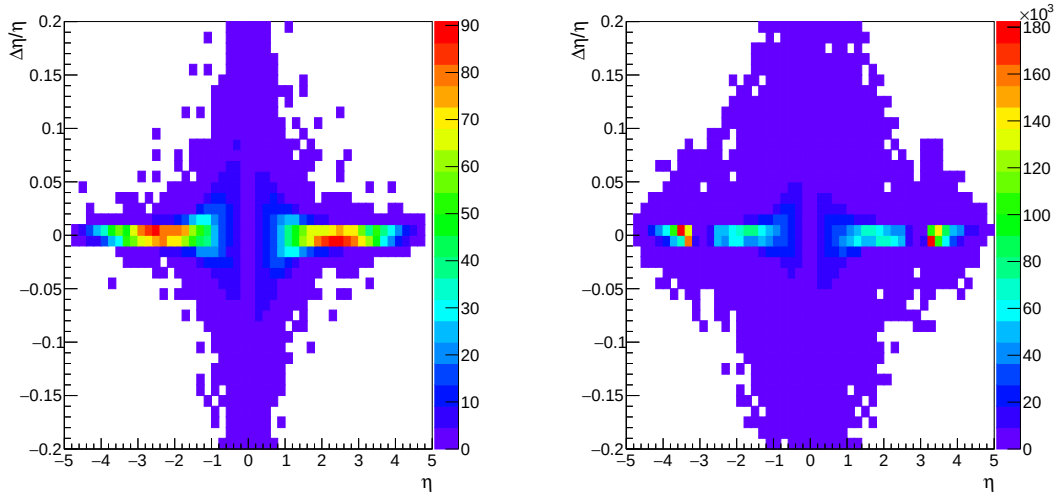


Figure 4.8: $\frac{\Delta\eta}{\eta}$ for the leading non b-jet, for Monte-Carlo events (left) and real data (right).

4.2.1 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 and Real data which

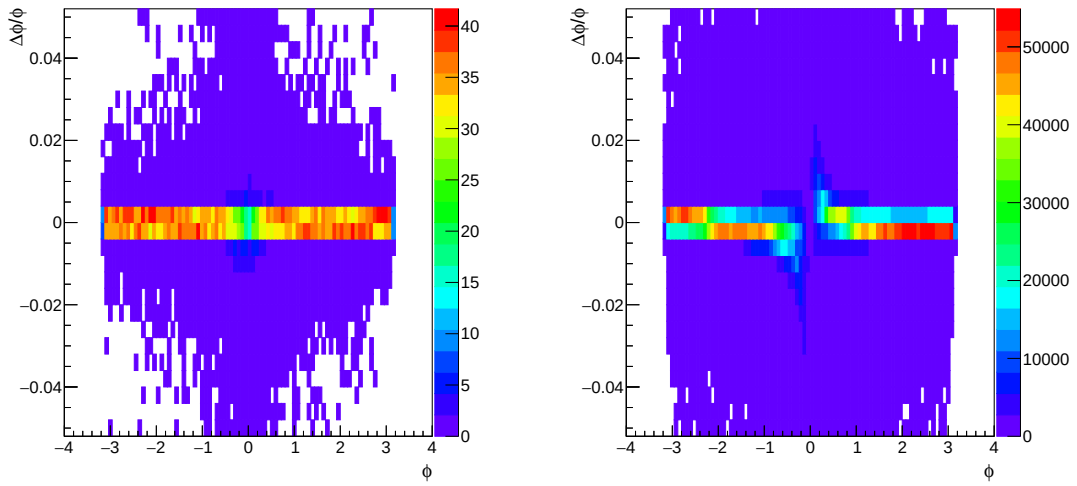


Figure 4.9: $\frac{\Delta\phi}{\phi}$ for the leading non *b*-jet, for Monte-Carlo events (left) and real data (right)

could be overcome with specific calibration of the jet objects reconstructed in the HLT.

4.3 Jet Tagging Efficiency

As covered in Section 2.5.3, the standard algorithm for 2016 physics analyses was chosen to be the 2016 MV2c10 algorithm. However, the HLT *b*-tagging algorithm uses the MV2c20 algorithm [33]. 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 being generated rather than recorded the *truth* nature of the jet object was known, and the result of the *b*-tagging algorithm can be compared to this truth label. This requirement for a truth label means only Monte-Carlo data can be used for this comparison.

In the analysis, an offline/HLT jet pair was formed using ΔR 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* working point. The efficiency plots in Figures 4.10, 4.11 and 4.12 show the fraction of these jets that were identified as *b*-jets by the HLT and offline tagging algorithms.

4.3.1 *b*-jet efficiency

For jets labelled as true *b*-jets, the tagging efficiency can be calculated and plotted with respect to topological variables of the jet objects.

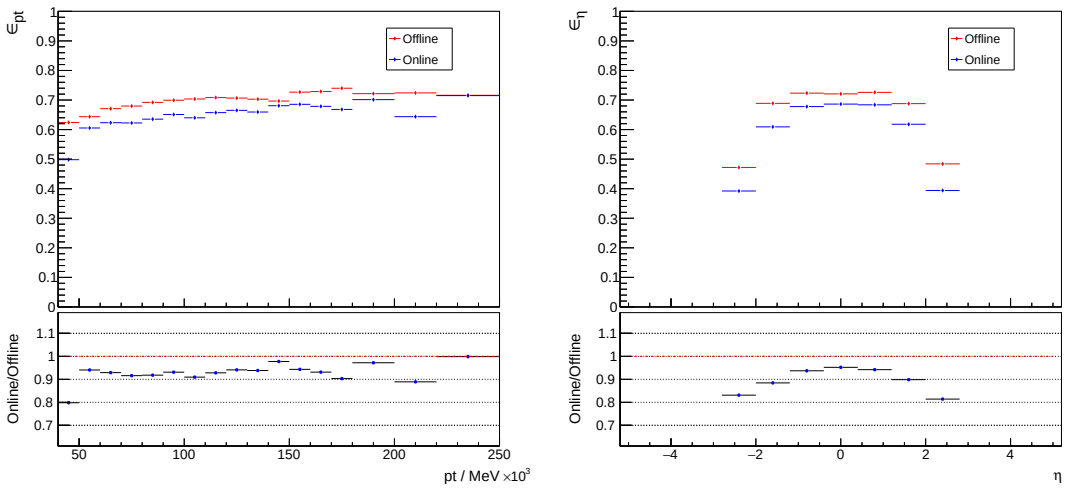


Figure 4.10: *b*-tagging efficiency for truth *b*-jets in Monte-Carlo data, plotted against jet p_{T} (left) and η (right).

4.3.2 *c*-jet efficiency

For *c*-jets and light-jets, plotting the same value gives the mistag rate for these jets in the detector.

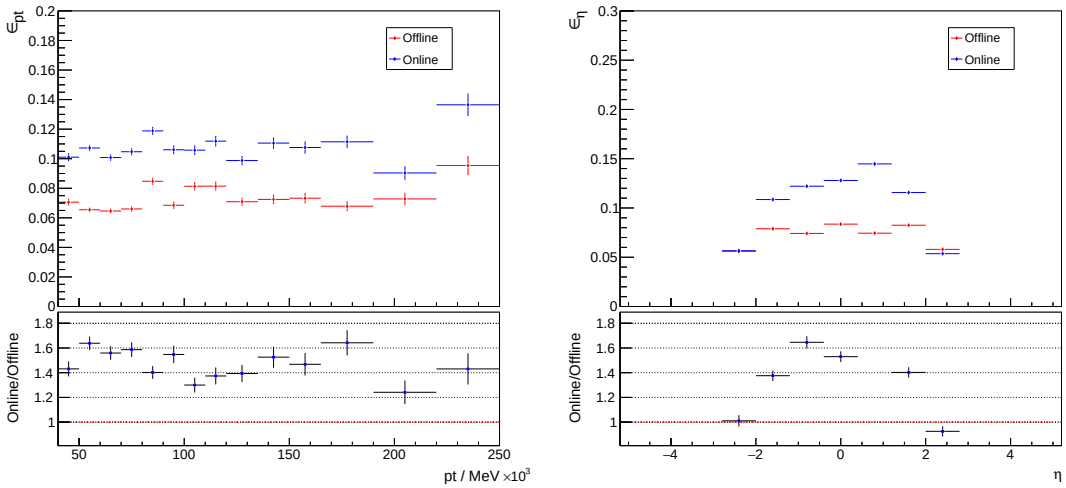


Figure 4.11: Mistag rate for truth *b*-jets in Monte-Carlo data, plotted against jet p_{T} (left) and η (right).

4.3.3 Light-jet efficiency

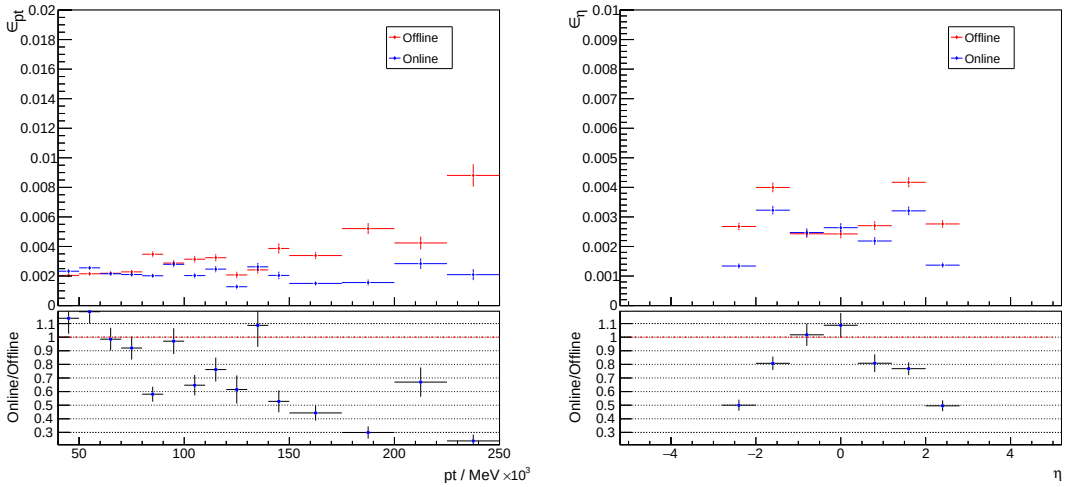


Figure 4.12: Mistag rate for truth light-jets in Monte-Carlo data, plotted against jet p_{T} (left) and η (right). Analysis is confined to the central region of the detector where b-tagging is operational.

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.

4.3.5 Comparison of HLT and offline tagging efficiencies

Primarily considering the p_{T} plots of efficiency, the HLT b-tagging is found to be around 5% less efficient than the offline b-tagging for jets with $p_{\mathrm{T}} > 50$GeV. This is a consistent direction of efficiency shift as found when comparing the 2016 MV2c10 and 2015 MV2c20 algorithms on the training $t\bar{t}$ sample, but of a larger magnitude. 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 $\sim 40\%$ increase is consistent with the expected shift from the optimised algorithm [44].

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 3.5 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 5.1, and the ratio of the events is shown in Figures 5.1 and 5.2.

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 4.10, 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 for two b-jets , which

Table 5.1: Cutflow for the *two-central* VBF $H \rightarrow b\bar{b}$ events as described in Section 3.5. 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
≥ 2 *loose* b-jets	503.552	467.146	2275760	2932620
≥ 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

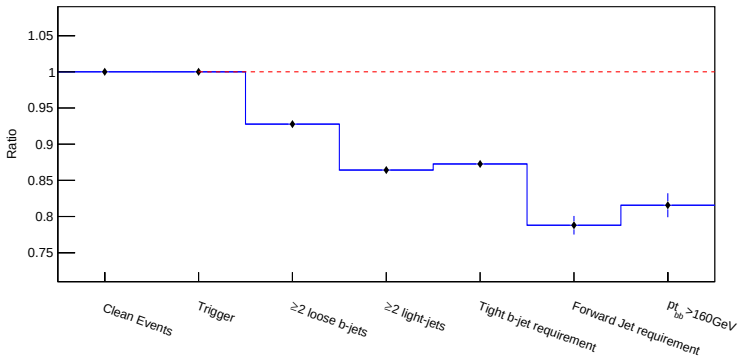


Figure 5.1: Ratio of the online event count over the offline event count for the Monte-Carlo

is lower than shown in the cutflow. As shown for the leading b-jetin Section 4.1, 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 4.7 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 $H \rightarrow b\bar{b}$ cuts.

5.1.2 Data

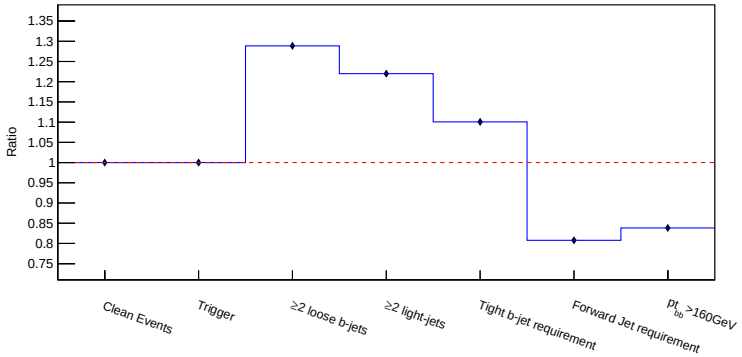


Figure 5.2: Ratio of the online event count over the offline event count for the real data

5.2 Specific Jet Feature Distributions

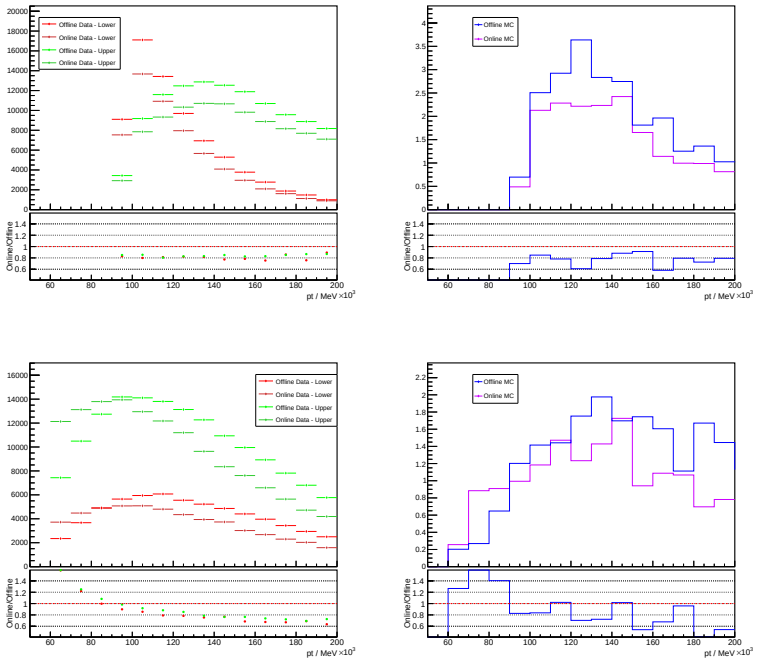
As the previous chapter showed both b-jets and non b-jets to be similar for online and offline objects, the kinematic properties of the jets that compose the VBF $H \rightarrow b\bar{b}$ event are shown to behave similarly.

5.2.0.1 p_{T}

5.2.0.2 η

5.3 BDT Input Variables

- M_{jj}
- $p_{\mathrm{T}\ jj}$
- $\cos\theta$
- $\Delta\eta_{jj}$
- $Max(\eta)$



- $\eta*$
- $min\Delta R(j_1)$
- $min\Delta R(j_1)$
- p_{T} balance
- $N_{TRK}(j_1)PV500$?
- $N_{TRK}(j_)PV500$?

5.4 Mbb Distribution

Prior paper suggests this is the 'final' plot, a shape comparison between BDT influenced control and signal regions of the Mbb distribution. A little confused as to exactly what we need here.

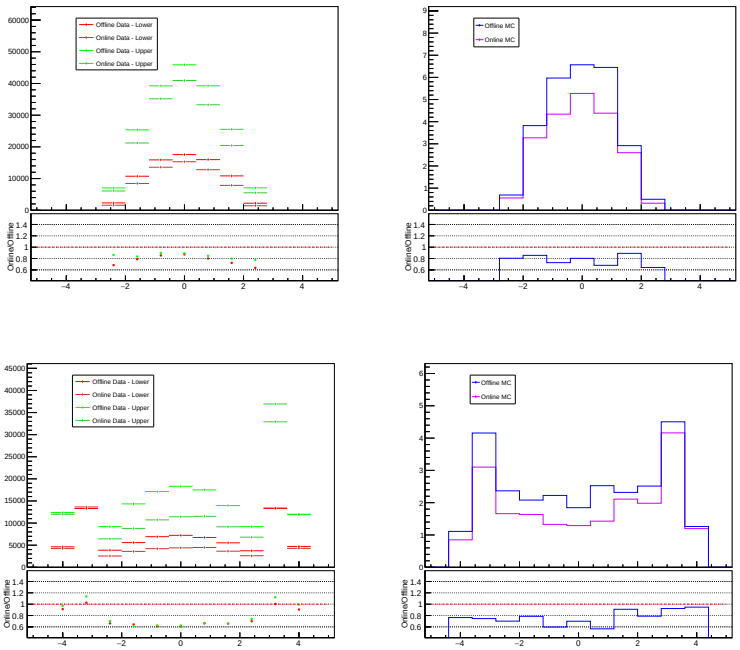


Figure 5.3:

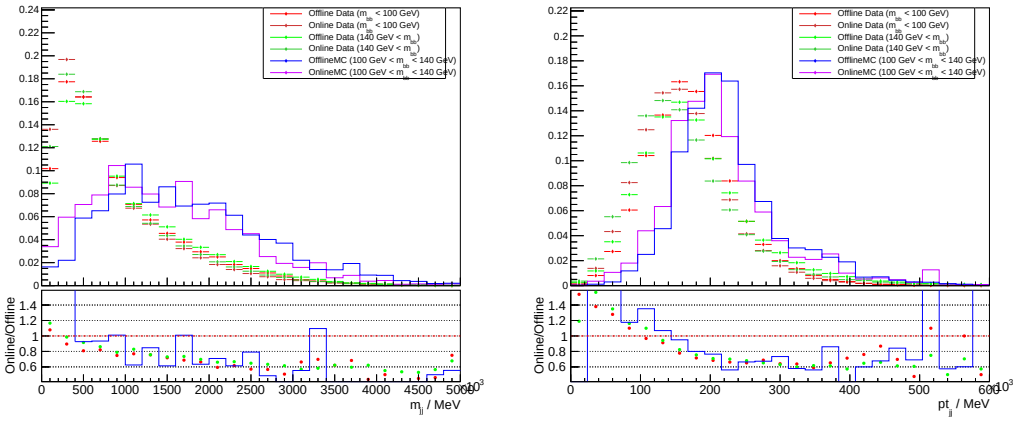


Figure 5.4:

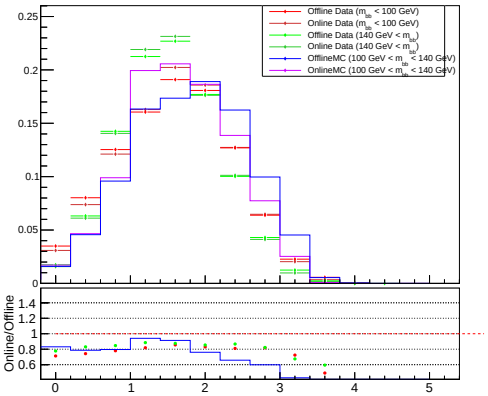


Figure 5.5:

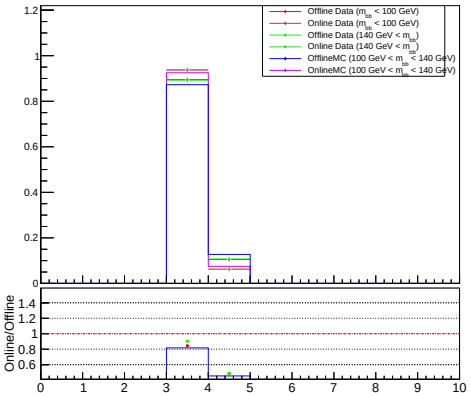


Figure 5.6:

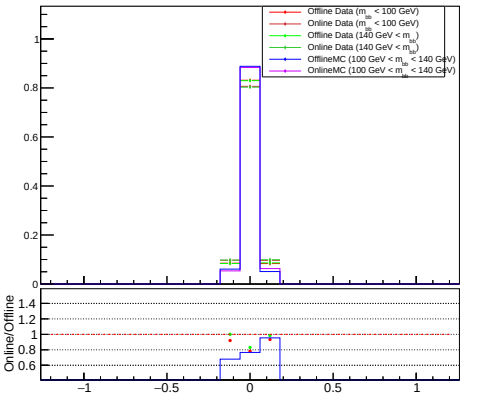


Figure 5.7:

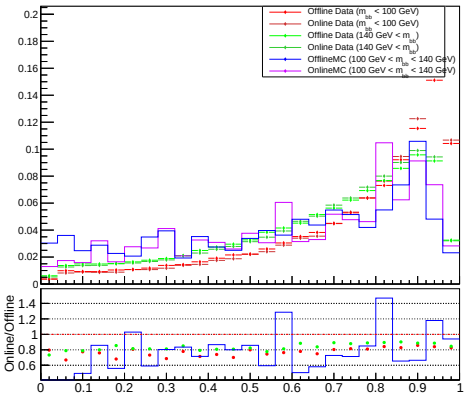


Figure 5.8:

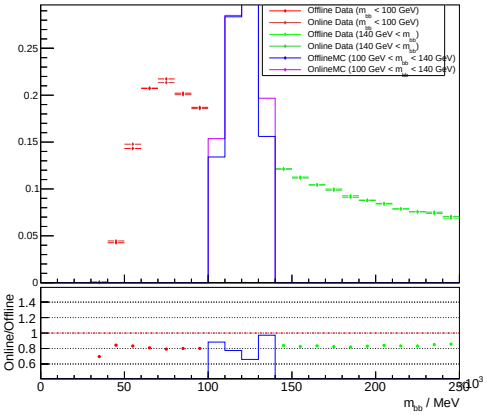


Figure 5.9:

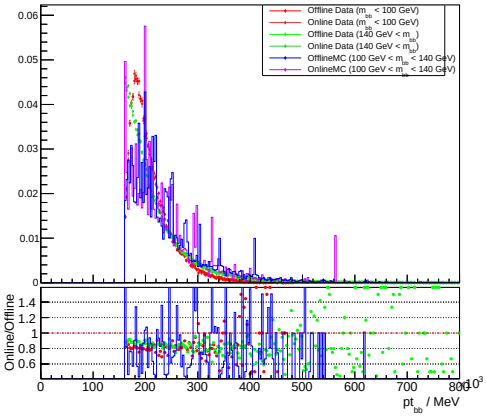


Figure 5.10:

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 $H \to b\bar{b}$ 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 [54]. Such structures are used throughout modern physics analyses at ATLAS [55].

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. [56]. 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.
$min\Delta R_{j1}$	Minimum (η, ϕ) separation between the leading VBF jet and the closest other jet.
$min\Delta R_{j2}$	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 [57].
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: $\frac{\boldsymbol{p}_{\mathrm{T}\boldsymbol{j_1}}+\boldsymbol{p}_{\mathrm{T}\boldsymbol{j_2}}+\boldsymbol{p}_{\mathrm{T}\boldsymbol{b_1}}+\boldsymbol{p}_{\mathrm{T}\boldsymbol{b_2}}}{p_{\mathrm{T}j1}+p_{\mathrm{T}j2}+p_{\mathrm{T}b1}+p_{\mathrm{T}b2}}$.
ΔM_{jj}	Difference in the largest invariant mass from all jet pairs and the invariant mass of the VBF jet pair

Bibliography

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

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

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

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

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

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

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

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

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

[10] 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].

[11] 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].

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

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

[14] 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.

[15] 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.

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

[17] 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].

[18] 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].

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

[20] 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].

[21] 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].

[22] 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].

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

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

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

[26] 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.

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

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

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

[30] 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.

[31] 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.

[32] 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.

[33] 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].

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

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

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

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

[38] 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.

[39] 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.

[40] 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.

[41] *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.

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

[43] *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.

[44] 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.

[45] 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.

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

[47] 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*,.

[48] 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.

[49] 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.

[50] 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.

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

[52] 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].

[53] 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].

[54] 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].

[55] 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.

[56] 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].

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

1