

Measurements of $H \rightarrow b\bar{b}$ decays and VH production

Thomas Charman

Supervisor: Dr. Jonathan Hays



Queen Mary University of London

Submitted in partial fulfillment of the requirements of the Degree of
Doctor of Philosophy February 8, 2021.

Contents

1	Introduction	5
2	Physics Theory	7
2.1	Historical Aside	11
2.2	Quantum Electrodynamics	12
2.3	Quantum Chromodynamics	14
2.4	Electroweak theory	15
2.5	The Brout-Englert-Higgs Mechanism	17
2.6	Higgs bosons at the LHC	19
3	The ATLAS Detector at the Large Hadron Collider	21
3.1	The Large Hadron Collider	21
3.2	The ATLAS Detector	26
	Tracking: The Inner Detector	28
	Pixel Detectors	31
	Semiconductor Tracker	31
	Transition Radiation Tracker	33
	Calorimeters	33
	Electromagnetic Calorimeter	35
	Hadronic Calorimeter	35
	Muon Spectrometers	36
	Trigger Systems	37
4	Machine Learning	39

4.1	Boosted Decision Trees	40
4.2	Neural Networks	41
4.3	Parametrised Neural Networks	47
5	Reconstruction and Selection	48
5.1	Athena and the CxAOD Framework	48
5.2	Leptons	51
Electrons	52	
Muons	52	
Taus	53	
5.3	Triggers	53
0-lepton channel triggers	54	
1/2-lepton channel triggers	55	
5.4	Jets	55
Topological Calorimeter Cluster anti- k_t jets	55	
b-tagging	56	
Pseudo-continuous b-tagging	58	
Truth tagging	59	
Hybrid tagging	61	
5.5	Missing Transverse Momentum	61
5.6	Overlap Removal	64
5.7	Final Selection	65
0-lepton channel selection	66	
1-lepton channel selection	67	
2-lepton channel selection	67	
6	Fit Models	69
6.1	Profile Likelihood Fits	69
6.2	VH(b,b) multi-variate discriminant fit	69
6.3	Di-jet mass fit	69
6.4	VZ Cross-check fit	69

7 Background Modelling	70
7.1 Categorisation into Analysis Regions	70
Top $e\mu$ control region	70
$\Delta R(b,b)$ Control Regions	70
7.2 Novel Modelling Techniques	70
Multi-dimensional re-weighting	70
Inferring missing information with parametrised neural networks . .	70
7.3 $Z + \text{jets}$ modelling	70
7.4 Modelling of other backgrounds	70
8 Results	72
8.1 VH(b,b) multi-variate discriminant fit results	72
8.2 Di-jet mass fit results	72
8.3 VZ Cross-check results	72
9 Conclusion	73
9.1 Future studies	73
Bibliography	74
Truth tagging for pseudo-continuous b-tagging	83

Chapter 1

Introduction

In 2012 the Higgs boson was discovered by the ATLAS and CMS collaborations at the Large Hadron Collider [1, 2]. It was said to form the last piece of the Standard Model of Particle Physics, a framework that describes three of the four fundamental forces of nature, described in more detail in Chapter 2. Despite apparent completeness after the Higgs discovery, it is known that the theory does not describe gravity, the fourth of the known fundamental forces of nature. The theory also has other shortcomings, it cannot explain the presence of dark matter [3–13] or a number of other observed phenomena [14–18]. So far the model has stood up to all experimental tests [19, 20] concerning its own predictions but there are still parameters of the model that have not been measured. Given the theory’s understood shortcomings, it is hoped that continued scrutiny of the models predictions will yield unexpected results, perhaps hinting at a new way forwards in terms of a theory that describes everything or simply exposing further gaps in our knowledge of the universe. For this reason it is more important than ever to study in detail the most recently discovered piece of the model, the Higgs boson.

This work focuses on studying a specific production mechanism and decay mode of the Higgs boson, specifically a vector boson associated Higgs boson decaying to two bottom quarks, denoted $VH(bb)$. This decay mode is of importance as it is currently the only decay mode of the Higgs decaying to quarks that has been observed [21]. A summary of the full spectrum of production mechanisms and decay modes of the

Higgs will be given in Chapter 2.

The study of this decay mode was carried out with the ATLAS detector, and made possible by the hard work of all members of the ATLAS collaboration. In Chapter 3 the detector is described in full.

Chapter 2

Physics Theory

The following chapter outlines the physics theory that informs and guides the experimental process of searching for new particles or making measurements at a particle collider. Not only are the physics theories described here useful in that context but they also provide an almost complete picture of the universe at certain scales. This chapter was written with the aid of notes taken at the annual STFC High Energy Physics Summer School, and with the aid of several books [22, 23], in which a more detailed description of the theories can be found.

The Standard Model of particle physics is a theoretical framework that describes all elementary particles and three of the fundamental forces of nature. Notably the only force that is not described by the theory is gravity. Particles described by the model are listed in table 2.1 with a white gap separating the matter particles (fermions) from the force carrying particles (bosons). Fermions, which make up solid matter obey Fermi-Dirac statistics [24, 25] whereas bosons obey Bose-Einstein statistics [26]. The Higgs boson is special in that as far as we know it does not carry a force in the conventional sense, instead it is responsible for giving fundamental particles mass, discussed in more detail in section 2.5. In the table of particles quarks (blue) and leptons (red) are ordered in columns by increasing mass, apart from the neutrinos, which are massless in the theory.

As mentioned the model describes forces as being mediated by certain particles, the photon (γ) mediates the electromagnetic force, particles experiencing electro-

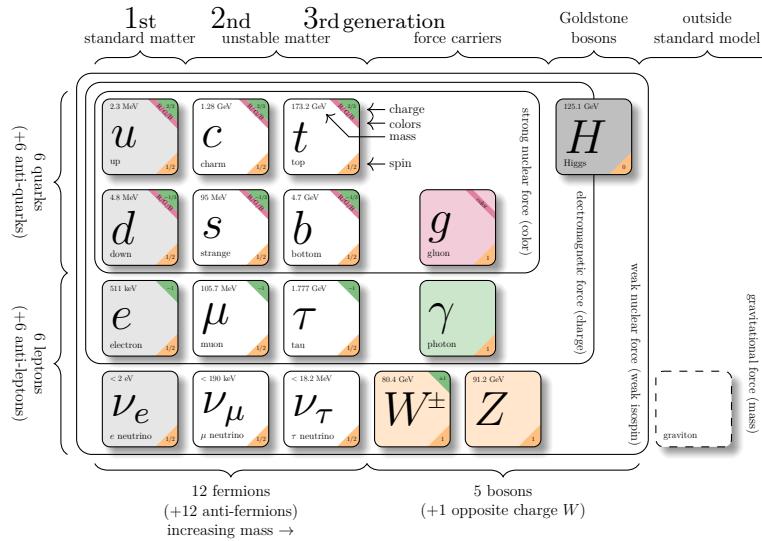


Table 2.1: The particles of the standard model with fermions displayed on the left and bosons on the right. Quarks are in blue, leptons in red, vector bosons in yellow and scalar bosons in green.

magnetic repulsion or attraction are described as exchanging photons. The strength and direction of the force experienced is proportional to the electromagnetic charge of particles involved. The Standard Model is a theory of quantum fields in which the strength of interactions between fields, or particles which are described as excitations in the fields, is parametrised by something known as a coupling constant. It is natural to assume that the strength of the interaction between photons and charged particles is related to the electromagnetic charge of the particles involved. Indeed this is the case consider the Coulomb force between two protons,

$$F = \frac{e^2}{4\pi\epsilon_0 r}, \quad (2.1)$$

where e is the elementary charge, ϵ_0 is the electric constant and r is the distance between the two protons in question and also the energy of a photon given by

$$E = \frac{hc}{\lambda}, \quad (2.2)$$

where h is Planck's constant, c is the speed of light and λ is the wavelength of the

photon. The value of the ratio of these two quantities

$$\alpha = \frac{e^2 \lambda}{4\pi\epsilon_0 r hc}, \quad (2.3)$$

known as the fine structure constant, is the coupling constant that describes the strength of the interactions between the photon field and fields of particles with electromagnetic charge. So as is now clear this coupling constant does indeed depend on the electromagnetic charge of the objects involved and so it is not in fact constant.

As well as the electromagnetic force the Standard Model describes the strong nuclear force and the weak nuclear force, shortened to just the strong and weak forces respectively. Like the electromagnetic force they too are mediated by the exchange of particles, the gluons (g) carry the strong force and the W^\pm and Z^0 bosons carry the weak force.

The charge associated with the strong force is known as colour which can take values that are mapped onto colours in the visible spectrum (red, green, blue) for ease of description. For each of these colours an anti-colour is also allowed (anti-red, anti-green, anti-blue). Unlike with the electromagnetic charge, particles with colour charge are not found freely in nature. Instead we find particles known as hadrons which are bound states of quarks and anti-quarks (e.g. the proton). The phenomenon of coloured particles being bound in such a manner is known as colour confinement, and the bound states are described by the quantum numbers isospin (I) and hypercharge (Y_c). It is commonly assumed that all free particles in nature are colour singlets e.g. for a hadron the state could be written as

$$\frac{(r\bar{r} + b\bar{b} + g\bar{g})}{\sqrt{3}}, \quad (2.4)$$

where r , b and g represent red, blue and green charges respectively. This phenomenon is known as quark confinement. Gluons carry colour and anti-colour indicating that there should be nine possible quantum mechanical states for the gluon given the available number of colour/anti-colour combinations, however when one

considers that the strong force is exclusively short range, and therefore that there should be no free gluons (disallowing colour singlet gluons) the number of possible states is reduced to eight. The state of a particle, as far as its description with respect to the strong force is concerned, is given by a vector which lives in a vector space, in which elements of the Lie group $SU(3)_C$ act as unitary operators, where the C denotes that the group is associated with the colour charge. The $SU(3)$ group is the group of 3×3 unitary matrices whose determinant is one. These correspond to the eight generators of $SU(3)$ where in general for a group $SU(N)$ the number of generators is given by $N^2 - 1$.

Describing the weak force requires introducing further quantum numbers weak isospin T and weak hypercharge Y_W . The state of a particle with regards to the weak force is given by a vector which lives in a vector space in which elements of $SU(2)_L \times U(1)_{Y_W}$ act as unitary operators where the L denotes that only particles in left-handed chiral states interact with the weak force¹. Left-handed fermions are represented as doublets in the theory with weak isospin $T = 1/2$ whilst right-handed fermions are singlets with weak isospin $T = 0$.

Along the way we have described particle states with respect to particular forces as vectors living in some vector space where the action of the element of a group has been as a unitary operator. If we are to describe a particle state taking into account the full model, the group whose elements should act as unitary operators on the particle state (the gauge group) is $SU(3)_C \times SU(2)_L \times U(1)_{Y_W}$. For each of the groups in the direct product we have established a (gauge) symmetry and therefore due to Noether's theorem [27] there should be an associated conserved quantity. The conserved quantities in this case are the electric charge, the weak hypercharge and isospin and the colour charge.

¹More specifically only left-handed chiral particles participate in weak charged current interactions.

2.1 Historical Aside

This section provides some historical context surrounding the Dirac equation which will later be used as the starting point in the discussion of quantum electrodynamics, which is the sector of the Standard Model that describes electromagnetic interactions.

In 1905 Albert Einstein first proposed the idea of special relativity [28]. The aim of the idea was to unify the then inconsistent theories of Maxwell's electromagnetism and Newtonian mechanics. The result of Einstein's work was a theory of motion which agreed with the predictions of Newtonian mechanics at velocities much smaller than the speed of light but whose predictions were accurate also at much higher velocities (for which Newtonian predictions fail). Arguably, the most far reaching consequence of special relativity is that it demands that any equation of motion must be invariant under Lorentz transformations, at least in terms of the formulation of new theories is concerned. Many physical phenomena predicted by special relativity could be considered of higher consequence in general, for example the phenomena of length contraction, time dilation, energy-mass equivalence and the universal speed limit (equal to the speed of light in vacuum), all of which are extensively scrutinised experimentally [29–35]. It is the Lorentz transformation however that should be kept in mind for the following discussion, the transformation may be written as

$$\begin{aligned} t' &= \gamma(t - vx/c^2) \\ x' &= \gamma(x - vt) \\ \text{with } \gamma &= \frac{1}{\sqrt{1 - v^2/c^2}}, \end{aligned} \tag{2.5}$$

in a single dimension of space x and one of time t where v represents the velocity of the system described by the primed coordinates relative to the unprimed coordinates and c is the speed of light in vacuum.

Twenty years after Einstein introduced the ideas of special relativity Erwin Schrödinger postulated new ideas regarding the motion of quantum mechanical systems [36].

Though he knew his new equation was not invariant under Lorentz transformations, and therefore incomplete, Schrödinger's formulation of quantum mechanics changed the way physicists thought about the universe for ever. His famous equation

$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{x}, t) = \left(\frac{-\hbar^2}{2m} \nabla^2 + V(\vec{x}, t) \right) \Psi(\vec{x}, t), \quad (2.6)$$

describes the states of particles as wave-functions Ψ which can only be interpreted in a probabilistic manner and contains Planck's constant the quantum of action. This work had many consequences including the quantisation of the values of measured observables (meaning they can only take discrete values) and the descriptions of particles as waves.

It was the aim of Paul Dirac to make the Schrödinger equation Lorentz invariant and thus provide a more complete description of quantum systems. Along the way he came to the realisation that in order for his equation to satisfy his needs the wave-function had to be replaced with a four component spinor (ψ) and the introduction of matrices known now as the Dirac matrices (labeled γ^μ with $\mu = 0, 1, 2, 4$) was required. Though not the form he originally wrote down Dirac's Lagrangian density takes the form

$$\mathcal{L}_{Dirac} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \quad (2.7)$$

where the repeated up and down indices are implicitly summed over (∂_μ represents the partial derivative taken with respect to a spatial coordinate $\mu = 1, 2, 3$ or time $\mu = 0$).

2.2 Quantum Electrodynamics

In order to take Dirac's Lagrangian (eq. 2.7) and turn it into something that appropriately describes quantum electrodynamics (QED), we should consider a $U(1)$

gauge transformation of the Dirac spinor and it's adjoint

$$\begin{aligned}\psi \rightarrow \psi' &= e^{i\alpha(x)}\psi, \\ \bar{\psi} \rightarrow \bar{\psi}' &= e^{-i\alpha(x)}\bar{\psi},\end{aligned}\tag{2.8}$$

with $\bar{\psi} \equiv \psi^\dagger \gamma^0$ and where $\alpha(x)$ is a local phase. Under this transformation the Lagrangian transforms as

$$\mathcal{L}_{Dirac} \rightarrow \mathcal{L}'_{Dirac} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \bar{\psi}\gamma^\mu \alpha(x)\psi\tag{2.9}$$

which is not equivalent to the original due to the factor resulting from the derivative of the transformed spinor. Instead let us change the derivative to the gauge covariant derivative

$$D_\mu = \partial_\mu + ieA_\mu\tag{2.10}$$

where we interpret A_μ as the photon field, with coupling constant e , parametrising the interaction strength. The field is also referred to as the electromagnetic gauge field since it arrives during the process of making the Lagrangian invariant under the $U(1)$ group, the gauge group of electromagnetism. Note that here what we have labeled e is nothing more than the fine structure constant previously denoted α in eq. 2.3. The transformation of the new field under the action of the gauge is defined as

$$A_\mu \rightarrow A'_\mu \equiv A_\mu - \frac{1}{e}\partial_\mu \alpha(x).\tag{2.11}$$

This means that the action of the gauge covariant derivative on the spinor transforms as

$$D_\mu \psi \rightarrow D'_\mu \psi' = e^{i\alpha(x)} D_\mu \psi\tag{2.12}$$

which means that the new Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi\tag{2.13}$$

is invariant under the action of the gauge as desired. What remains in order to write a description of QED is to write down a kinetic term for the photon field. An appropriately gauge and Lorentz invariant term is

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (2.14)$$

where the electromagnetic tensor is defined as

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (2.15)$$

Putting everything together we can define the Lagrangian for QED as

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (2.16)$$

2.3 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the theory of the strong force. Its mathematical formulation is similar to that of QED. Except the gauge group for QCD is $SU(3)_C$ where the C denotes that the force is associated with colour charge. As previously discussed the eight generators of the group are associated with the eight gluons of the Standard Model. The generators are present in the form of the transformation of a fermion field under an element of $SU(3)$

$$\psi \rightarrow \psi' = \exp\left(i\alpha_a(x) \cdot \frac{\lambda_a}{2}\right)\psi, \quad (2.17)$$

where the λ_a are the Gell-Mann matrices, generators of $SU(3)$. A key difference between the strong force and the other forces described by the Standard Model is that it increases in strength with range. This property leads to a phenomena known as quark confinement which has been discussed previously. Quark confinement is the reason for many of the complications that arise when trying to detect certain particles in a particle detector such as ATLAS. Specifically, quarks that are produced

in collisions undergo a process called hadronisation whereby they transition from their coloured states to colour singlets. Excess energy present in this process results in the creation of lots of different states, some which decay to leptons, with the overall process producing a roughly conical shower of particles known as a jet.

2.4 Electroweak theory

The Glashow-Salam-Weinberg model of electroweak interactions [37–39] describes the weak force and electromagnetism as a quantum field theory, which is gauge invariant under transformations that are elements of $SU(2)_L \times U(1)_{Y_W}$. As previously mentioned the L and Y_W subscripts denote that the gauge groups in the direct product that are associated with left-handed chiral particles and weak hypercharge respectively. The association with weak hypercharge distinguishes this $U(1)$ group with the $U(1)$ group from QED. The transformation of the fermion fields under $SU(2)$ is given by

$$\psi \rightarrow \psi' = \exp\left(i\vec{\alpha}(x) \cdot \frac{\vec{\sigma}}{2}\right)\psi, \quad (2.18)$$

where $\vec{\sigma}$ is a vector of the Pauli matrices σ_i with $i = 1, 2, 3$, a familiar representation of $SU(2)$ generators. Constructing a gauge covariant derivative for the full transformation under $SU(2) \times U(1)$ requires the addition of new fields analogous to the photon field from QED, the new derivative takes the form

$$D_\mu = \partial_\mu - i\frac{g_1}{2}Y_W B_\mu - i\frac{g_2}{2}\sigma_i W_\mu^i, \quad (2.19)$$

where coupling constants g_1 and g_2 parametrise the strength of interactions with each field. The index i runs over the three Pauli matrices and three new fields W_μ^i with $i = 1, 2, 3$ which are associated with the $SU(2)$ gauge. The B_μ field is associated with the $U(1)_{Y_W}$ gauge and is obtained in the same way as the photon field in QED but is given a new symbol as it is *not* the photon field.

In fact none of the fields added here are the physical fields that we have access to in nature associated with the electromagnetic force or the weak currents. In order

to obtain the physical fields for the weak charged current one can simply take the linear superposition

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \quad (2.20)$$

In order to recover the photon field and reveal the field for the weak neutral current the idea of weak mixing must be introduced. Weak mixing was introduced to theory after the discovery of parity violation [40]. Parity is equivalent to chirality for massless particles, however for particles with mass a Lorentz boost can always appear to flip the chirality of the particles state whereas parity is a fundamental property of a particle. A fermion field with left or right handed chirality can be obtained by multiplication with one of two corresponding projection operators defined as

$$\begin{aligned} P_L &= (1 - \gamma^5)^2/2, \\ P_R &= (1 + \gamma^5)^2/2, \end{aligned} \quad (2.21)$$

with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, where γ^μ are the Dirac matrices, like so

$$\begin{aligned} \psi_L &= P_L\psi, \\ \psi_R &= P_R\psi. \end{aligned} \quad (2.22)$$

It is known that the weak neutral current and indeed the electromagnetic force both interact with particles of left and right handed chirality. Spontaneous symmetry breaking, theorised to have occurred due to an electroweak phase transition in the early universe, has the effect of rotating the plane defined by the B_μ and W_μ^3 fields into the physical fields we see in nature today. The mixing of the fields due to this rotation takes the form

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \quad (2.23)$$

where θ_W the Weinberg angle parametrises the amount of mixing. This picture shows that unification of QED and a description of the weak force has been achieved.

Although it at first seems like the $U(1)_{EM}$ gauge group is not present in $SU(2)_L \times U(1)_{Y_W}$ gauge group of electroweak theory it has been shown that the QED gauge symmetry is recovered by spontaneous symmetry breaking. Also the Y_W subscript in the gauge group represents weak hypercharge which is related to electric charge Q by the following relationship

$$Y_W = 2(Q - T^3), \quad (2.24)$$

where T^3 is the third component of isospin, the component that is conserved.

The particles associated with the weak neutral and charged currents are observed to have masses in nature [41–44] therefore one would naively like to write mass terms of the form

$$\mathcal{L}_{mass} \propto M_B^2 B^\mu B_\mu \quad (2.25)$$

$$+ M_W^2 W_a^\mu W_a^\mu. \quad (2.26)$$

The above mass terms are however not gauge invariant therefore another solution is required, one which will be discussed in the next section.

2.5 The Brout-Englert-Higgs Mechanism

The Brout-Englert-Higgs mechanism was made complete almost simultaneously by R. Brout and F. Englert [45], P. Higgs [46] and, G. Guralnik, C. R. Hagen and T. Kibble [47]. The underlying mechanism was proposed prior to this work by P. Anderson [48], though this initial theory was not relativistic invariant. It was initially proposed as a means to give the vector bosons mass terms that were gauge invariant. The theory predicts a complex scalar field (the Higgs field) that undergoes spontaneous symmetry breaking. Interactions with this field are predicted to be mediated by a massive spin-1 scalar particle that is now known to be the Higgs boson. This particle also gives mass to the fermions via a different mechanism. In

general spontaneous symmetry breaking is a process by which a symmetry breaks once conditions meet some threshold. An example of this is a hot sphere of ferromagnetic material whose spins are isotropically oriented. As the sphere cools the ferromagnetic property of the material will align the spins. In the hot scenario the sphere had symmetry in all spatial directions, by this it is meant that the changes to the sphere's orientation were indistinguishable. Once the spins have aligned however this is no longer the case, the fact that the spins point in a specific direction means that direction is special and so some of the symmetry was spontaneously broken. It can be noted though that a preserved symmetry still exists as rotations about the axis defined by the direction of the spins would leave the sphere invariant. In the Standard Model the symmetry that breaks is that of the complex scalar Higgs field. Consider a Lagrangian involving the field ϕ of the form

$$\mathcal{L} = T - V(\phi) = \partial_\mu \phi^\dagger \partial^\mu \phi - \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (2.27)$$

$$\text{with } \phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2). \quad (2.28)$$

Invariance under global phase transformations of the form $\phi \rightarrow e^{i\theta}\phi$ depends on the parameters of the potential μ and λ . Figure 2.1 shows two sketches of the potential for the scenarios where $\mu^2 > 0$, $\lambda < 0$ (left) and $\mu^2 < 0$, $\lambda < 0$ (right). To suggest

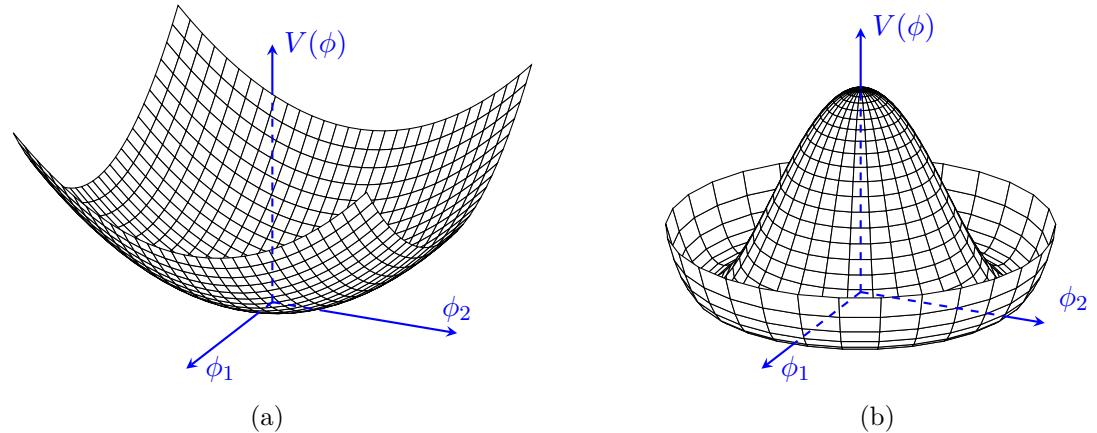


Figure 2.1: The Higgs potential in its fully and broken symmetric forms.

that in our universe this symmetry is spontaneously broken is to suggest that the values of these parameters evolved over time from the full to the broken state. This

ends up leading to masses for the vector bosons that are dependent on μ^2 .

2.6 Higgs bosons at the LHC

Higgs bosons are produced at the LHC in a number of different ways, the four most common of which are shown in figure 2.2. The prevalence of these processes

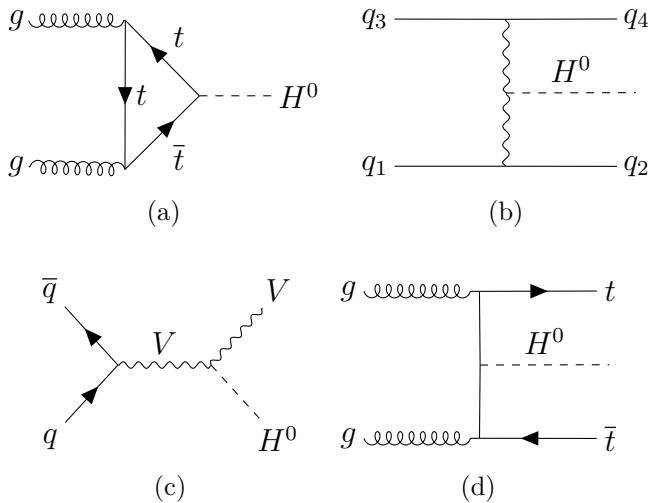


Figure 2.2: The four most common Higgs boson production methods from proton-proton collisions at the LHC.

with respect to the centre of mass energy of the proton-proton collision is shown in figure 2.3 (a). It can be seen the gluon-gluon fusion (fig 2.2 a) is by far the dominate contributor occurring over an order of magnitude more than the next highest process which is quark associated production (fig 2.2 b). The next highest production channel with respect to cross section is vector boson associated (fig 2.2 c) which will be the focus of the rest of this report. Finally top quark associated production (fig 2.2 d) has the smallest cross section of these processes. The Higgs boson is predicted by the Standard Model to decay in a number of different ways depending on its mass, a free parameter of the model. In figure 2.3 (b) the branching ratios of the Higgs can be seen, plotted with respect to Higgs mass. The decay that will be focused on for the rest of this report is $H \rightarrow b\bar{b}$. Given that the focus here is on vector boson associated production of a Higgs boson it is also important to consider the decay of the vector boson. Three possible scenarios are represented in

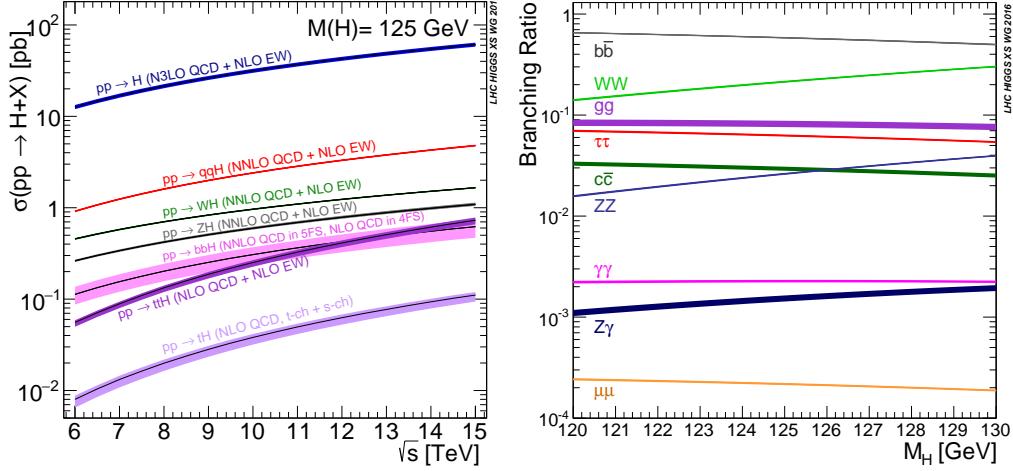


Figure 2.3: Higgs production cross-sections (left), and branching ratios (right) for a range of centre of meass energies and Higgs boson masses respectively [49].

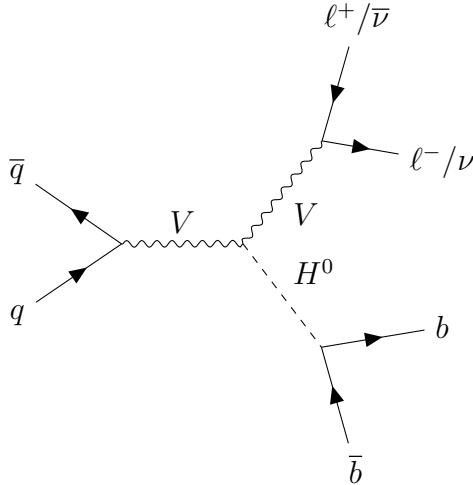


Figure 2.4: A diagram showing a Higgs boson (decaying to a pair of b quarks) produced in association with a vector boson (decaying to 0, 1, or 2 charged leptons denoted $\ell^{+/-}$).

figure 2.4, namely the situations where the vector boson decays to 1, 2 or 3 charged leptons and the appropriate number of neutrinos. It is in fact these leptonic decay modes that motivate the reason for studying this production mechanism as opposed to one of the more common ones. The issue with looking at the other production modes is that very large QCD generated backgrounds are present due to initial state radiation. Whilst these backgrounds are also present when looking at the vector associated channel they can be partially suppressed by triggering on a lepton.

Chapter 3

The ATLAS Detector at the Large Hadron Collider

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [50] is a large circular machine located 100 m underground straddling the Swiss-French border at the European Organisation for Nuclear Research (CERN). The LHC accelerates and collides protons and other charged particles. It has a diameter of 27 km and resides in a tunnel which was originally excavated for the Large Electron-Positron Collider [51] experiment. During its construction the tunnel was the largest civil engineering project in Europe to date. Today there are many physics experiments that take place at CERN, some of which are marked in figure 3.1. There are currently seven experiments that record data from the collisions at the LHC: ATLAS [52], CMS [53], LHCb [54], ALICE [55], MoEDAL [56], TOTEM [57] and LHCf [58].

The Lorentz force is fundamental to the LHC's accelerator technologies and detectors. Expressed as

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}), \quad (3.1)$$

it is clear that the force due to an electric field \vec{E} on a particle with charge q acts in the direction of the velocity of the field whereas the force due to a magnetic field \vec{B}

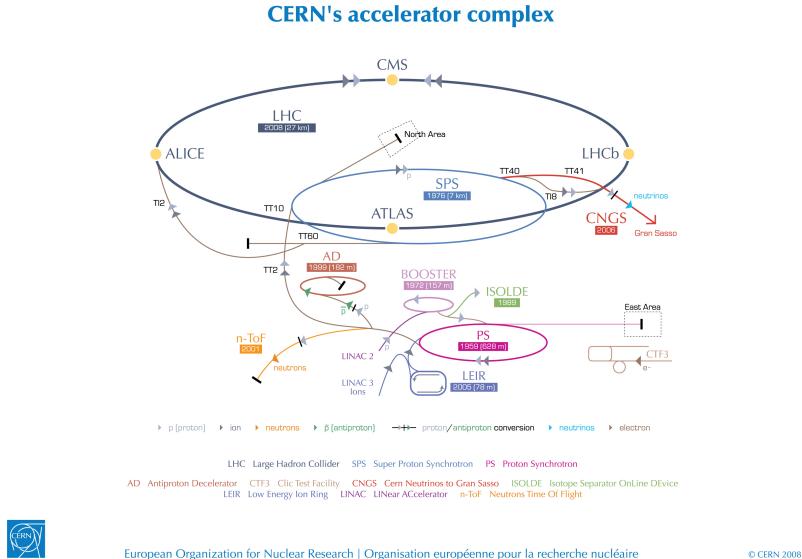


Figure 3.1: The CERN accelerator complex [59].

acts perpendicular to both the field and the velocity of the particle \vec{v} . It is therefore clear that an electric field may be used to accelerate and give energy to a charged particle whereas a magnetic field will alter the trajectory of a particle whilst keeping its energy constant.

The LHC is a synchrotron, an accelerator that uses magnets in a dipole configuration, such as in figure 3.2, to bend the path of charged particles into conformity with its circular shape. It is apparent from studying the figure that counter-rotating beams of same sign charged particles will require two sets of dipole magnets in order to rotate in opposite directions around the same ring. This is one disadvantage of a proton-proton collider with respect to a proton-anti-proton collider such as the Tevatron [60] which can use the same magnets for both beams. On the other hand, the proton-proton collider is able to separate the beams after they have been brought together to collide with a single dipole. The bending magnets of a synchrotron are designed to ramp up their magnetic field in synchronisation with the kinetic energy of the accelerated particles, allowing higher energies to be achieved before the beam is lost. The LHC can accelerate each beam to an energy of 6.5 TeV leading to collisions with a centre of mass energy of $\sqrt{s} = 13$ TeV, although the design energy of the LHC is $\sqrt{s} = 14$ TeV. The Tevatron held the previous record for centre of mass energy of collisions of $\sqrt{s} = 2$ TeV. The LHC has 1232 dipole magnets [50] which

are made of copper-clad niobium-titanium cables, a superconducting material whose electrical resistance falls to zero below 10 K. In order to maintain super-conductivity a cryogenic system using liquid helium is employed to cool the magnets. The higher the velocity of a charged particle, and the tighter the desired bending radius, the larger the magnetic field required to perform the bending. The large size of the LHC and the choice of superconducting magnet technologies are both informed by the aim to accelerate protons to the highest energy, and therefore velocity, possible.

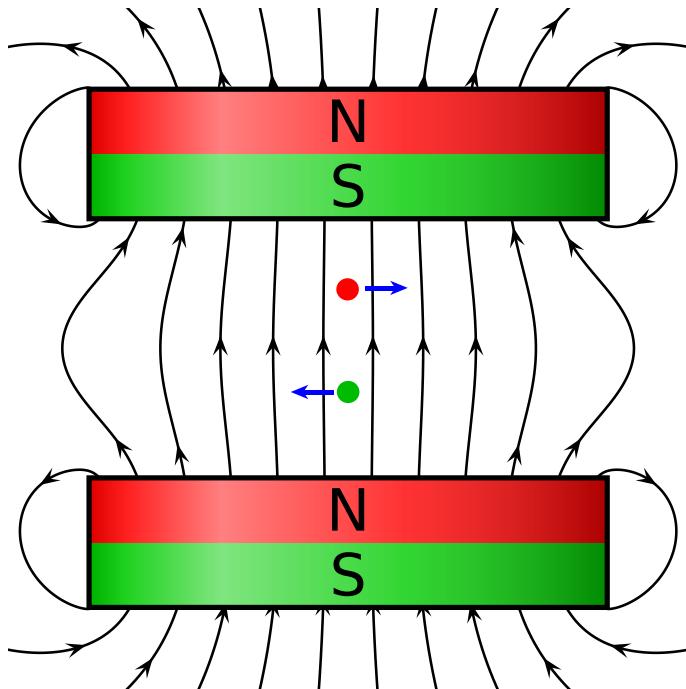


Figure 3.2: A representation of a pair of idealised cylindrical magnets in a dipole configuration. Two positively charged particles are shown as circles, the red particle is traveling out of the page, the green particle is traveling into the page. The forces experienced by each particle due to the magnetic field are shown as blue arrows.

The force which accelerates the particles is provided by radio-frequency cavities such as in figure 3.3, of which the LHC has 16 [50] The electric field in the radio-frequency cavity forms a standing wave, the separation between bunches of particles to be accelerated must be matched to the frequency of this wave. Protons in the LHC are accelerated in bunches and in vacuum, this to increase the likelihood of collisions and mitigate loss of energy and scattering effects due to interactions with air molecules. These two factors lead to the occurrence of space charge which causes

an increase in the emittance of the beam, where the emittance is defined as the total area that the beam occupies in its beam-pipe. The greater the energy of the particles the more they can overcome increase in emittance due to space charge. Increased emittance is especially problematic in circular accelerators where periodic effects can quickly lead to the loss of beam. For these reasons it would be very challenging to accelerate protons from rest in a synchrotron, the starting point for the protons of the LHC is therefore a linear accelerator called Linac2 [?] which is used to overcome space charge effects before the protons move on to a series of synchrotrons as seen in figure ??.

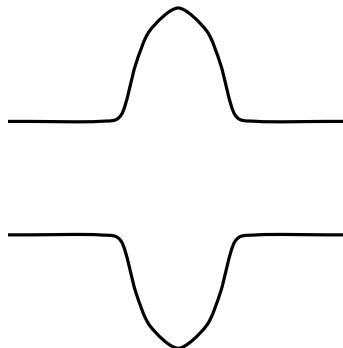


Figure 3.3: A representation of a radio-frequency cavity where the electric field lines are shown in black.

Even a beam with its emittance under control would still be lost from the accelerator if only dipole magnets were used to control its path. Magnets in a quadrupole configuration as in figure 3.4 are used to focus the beam and keep it in the beam-pipe. The quadrupoles behave such that particles feel a force that increases with the distance from the centre of the beam leading to simple harmonic motion of individual particles in a bunch. The LHC has a series of 24 quadrupole magnets each for focusing in the horizontal and vertical directions [50] as well as higher multipolarity configurations; sextupole, octupole, decapole and dodecapole which are used to correct imperfections in the fields of other magnets.

The statistical nature of particle physics analyses means that larger datasets (more recorded collisions) increase the sensitivity of searches and measurements. Constraints on the number of years the LHC is able to run for mean that the best way to record more collisions is to collide more particles per second. A quantity

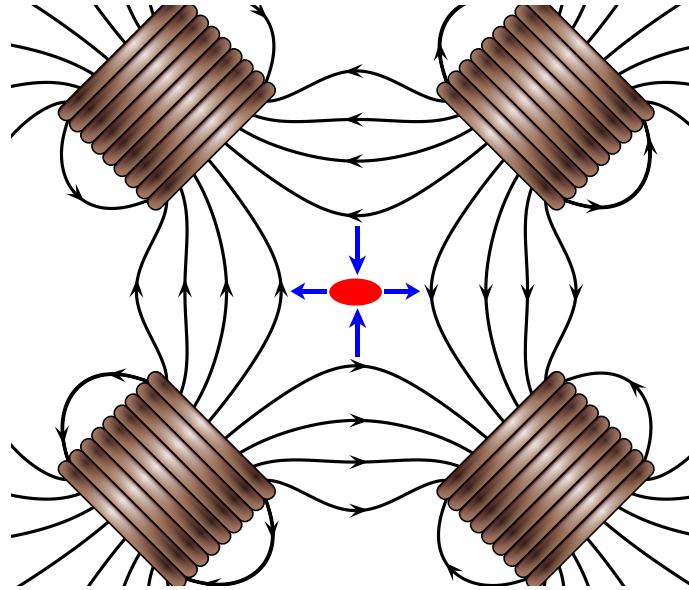


Figure 3.4: Representation of an idealised set of coil magnets in a quadrupole configuration with a proton beamspot shown as a red ellipse. The proton beam is drawn coming out of the page (*watch out!*), magnetic field lines are drawn in black and the forces acting on each bunch of protons are drawn as blue arrows.

known as the luminosity is often used to describe how much data is available for an analysis, it is written as

$$L = \frac{1}{\sigma} \frac{dN}{dt}, \quad (3.2)$$

where σ is the cross-section, a volume within which particles must pass by one another in order to interact, and N is the number of events recorded in a period of time t . For luminosity at the LHC N can be expressed as

$$N = n_{bp} n_1 n_2 \nu_r, \quad (3.3)$$

where n_{bp} is the number of colliding bunch pairs, n_1 and n_2 are the number of protons in each beam and ν_r is the frequency with which the beams rotate around the LHC's circumference. The number of particles in the beams is limited by space charge. The number of bunches is limited by the frequency that the radio-frequency cavities can operate at. The revolution frequency is limited by the strength of the dipole magnets and the circumference of the accelerator ring. Increasing luminosity by reducing the cross-section amounts to reducing the beam widths which is limited by the emittance of the beam. The LHC has already exceeded its design luminosity

providing physicists with more data to analyse than expected and plans are well underway for the upgrade to a High-Luminosity LHC (HL-LHC) [61].

3.2 The ATLAS Detector

The ATLAS detector [62] resides at a location on the LHC ring called Point 1, its full name is A Toroidal LHC ApparatuS. A diagram of the detector is shown in figure 3.5. ATLAS is considered to be a general purpose particle detector and has a wide physics program including: Higgs boson physics, top quark physics, searches for Supersymmetry and exotic states, probes of CP violation in b-quarks and light states and heavy ion physics. The detector itself is very large in size, spanning a width of 25 m and a length of 44 m and weighs 7000 tonnes which is comparable to the weight of the wrought iron content of the Eiffel tower [63].

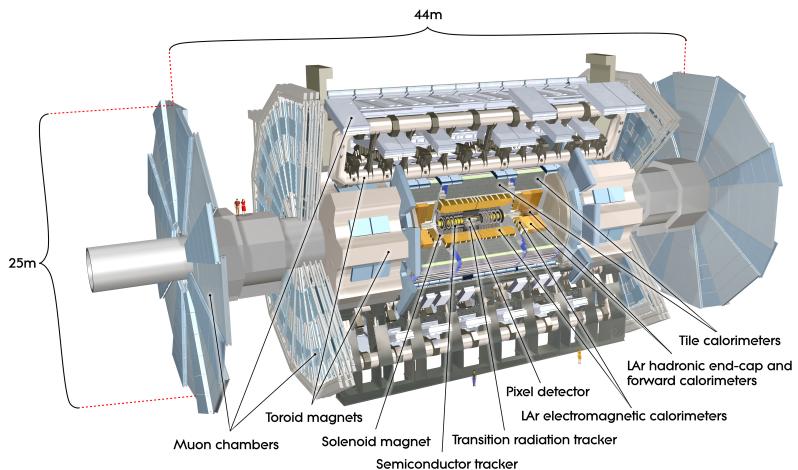


Figure 3.5: Computer generated image of the whole ATLAS detector with the major sub detectors labeled [64].

Due to the composite nature of the proton, the decay products of collisions are extremely numerous. Additionally when two bunches of protons cross there is the chance that more than one hard scattering event occurs and softer glancing collisions are also a possibility. The number of hard scattering events in a given bunch crossing is known as the pile-up of the collision and is often denoted with the symbol μ . As can be seen by inspecting equations 3.3 and 3.2 increasing the luminosity will often

cause a higher pile-up environment in the detector. The variety of decay products of each collision necessitate the use of specialised sub-detectors in order to accurately measure the output of collisions. There are dedicated sub-systems in ATLAS with the purpose of measuring properties of specific classes of particles. Ultimately in all cases a digital electrical signal is the desired output of a sub-detector which usually originates as an analogue signal. It is interesting to note that despite the many charges that are associated with the forces of nature discussed in chapter 2 the only one that we can directly measure is electric charge.

The ATLAS sub-systems are located in either the barrel of the detector or one of the end-caps. These two areas have a different geometry, and due to their relative positions with respect to the beam-pipe are exposed to different amounts of radiation, so the design of a sub-system in the barrel will differ from that of a sub-system measuring the same quantities in the end-cap. The details that follow are based on the ATLAS technical design report volumes [65, 66] unless another citation is present. Before detailing individual components it is important to detail certain properties of the detector relevant to all sub-systems. The coordinate system used to describe the ATLAS detector is known as right-handed. Three orthogonal axes (x, y, z) are used to describe the 3D space of the detector. The x-axis points towards the centre of the LHC ring, the y-axis points upwards and the z-axis points along the LHC beam pipe y-axis. The three axes meet at the interaction point which is the nominal position where bunches cross, located in the centre of the detector. Cylindrical coordinates (r, ϕ) are also often used to describe the physical features of the detector and phenomena caused by interactions in the detector that shall be referred to as analysis objects. Their definitions are that ϕ is the azimuthal angle in the x-y plane (transverse) around the beam pipe and r is the distance from the interaction point. A relativistic invariant analogy to the zenith angle of an object measured from the z-axis is pseudo-rapidity $\eta = -\ln(\tan(\theta/2))$ where θ is the ordinary zenith angle measured from the same axis.

By inspecting equation 3.1 it is clear that neutral particles for which $q = 0$ do not experience the Lorentz force. However for charged particles, as previously discussed,

a magnetic field alters the trajectory of the particle whilst preserving its energy. A large portion of the ATLAS detector is immersed in magnetic fields created by the magnet systems, and so this deflection phenomenon is present and exploited when making measurements of charged particles. There are four magnet systems in ATLAS the solenoid, the barrel toroid, and two end-cap toroids. The solenoid surrounds the inner detector whilst the toroid systems surround the muon chambers. Figure 3.6 shows a heat map of the magnetic field strengths within ATLAS, the image is from an article detailing the superconducting magnet system [67]. The magnet systems store a total energy of 1.6 GJ and produce fields of a combined volume of approximately $12 \times 10^4 \text{ m}^3$.

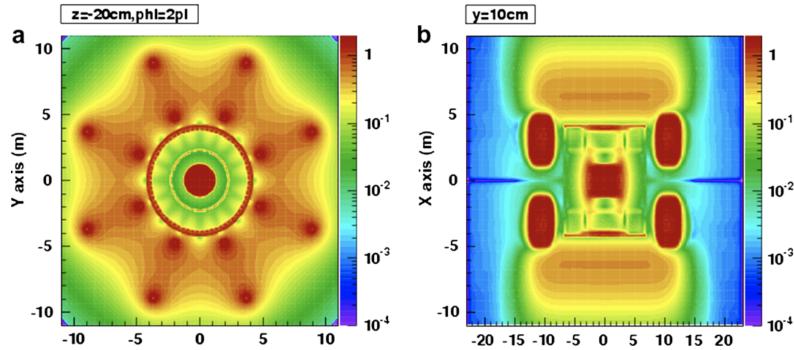


Figure 3.6: ATLAS magnetic field profile, showing a transverse cross-section in the centre of the detector (a), and a longitudinal section (b) [67].

Tracking: The Inner Detector

The Inner Detector (ID) is comprised of a number of different tracking detector sub-systems; the pixel detectors, the semiconductor tracker (SCT) and a transition radiation tracker (TRT) as seen in figure 3.9. It covers a volume corresponding with the total ϕ angle. The pixel detectors and SCT cover the range $|\eta| < 2.5$ and the TRT covers $|\eta| < 2.0$. The job of the ID is to track the propagation of charged particles through the detector. This is achieved by measuring a sequence of hits for each charged particle that propagates through its volume. The sequence of hits describing the trajectory of a given charged particle must be disentangled from other hits in the detector, track-finding algorithms are applied to this end. Using these

tracks we can measure; the direction of the particle, the sign of the electric charge of the particle, the rate of energy loss of the particle with respect to its distance traveled and by constructing the sagitta the transverse momentum of the particle (denoted p_T). An example construction of the sagitta is shown in figure 3.7 and its relation to p_T is

$$S = \frac{qL^2B}{8p_T}, \quad (3.4)$$

where L is the distance between the first and last hits in the track, q is the charge of the particle, and B is the strength of the magnetic field, in this case the field produced by the solenoid magnet system. The system is made of a single layer

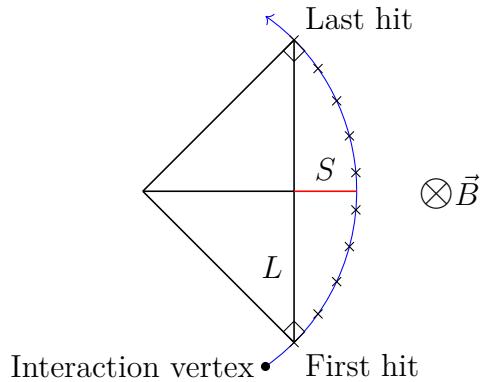


Figure 3.7: A diagram showing the geometric construction of the sagitta (S) of a track. The track is comprised of a sequence of hits marked by crosses and the “lever-arm” distance between the first and last hit in the track is marked L . Charged tracks in a magnetic field pointing into the page, as shown, form arcs of circles. The two lines marked as normal to the track are radii of the circle to which this track’s arc belongs.

coil with an inner diameter of 2.46 m and produces 2 T field in the axial direction with respect to the beam-pipe. Using multiple tracks interaction vertices can also be reconstructed, the vertex which comes from the highest energy collision in a given event is known as the primary vertex. Vertices coming from pile-up and glancing collisions of partons also need to be reconstructed in order to separate detector response due to these collisions from that of the primary vertex. So-called impact parameters d_0 and z_0 are used to identify secondary and pile-up vertices, their construction is shown in figure 3.8. These parameters are also used to perform identification of particles within the detector such as b -quarks and τ s. Tracks are

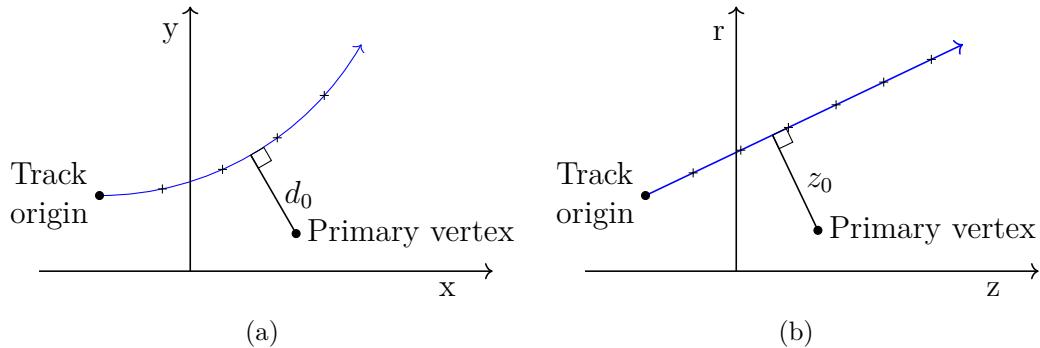


Figure 3.8: A diagram showing the geometric construction of the impact parameters d_0 and z_0 .

also used to calculate the calorimeter impact point and in general match activity in outer regions of the detector to an interaction vertex.

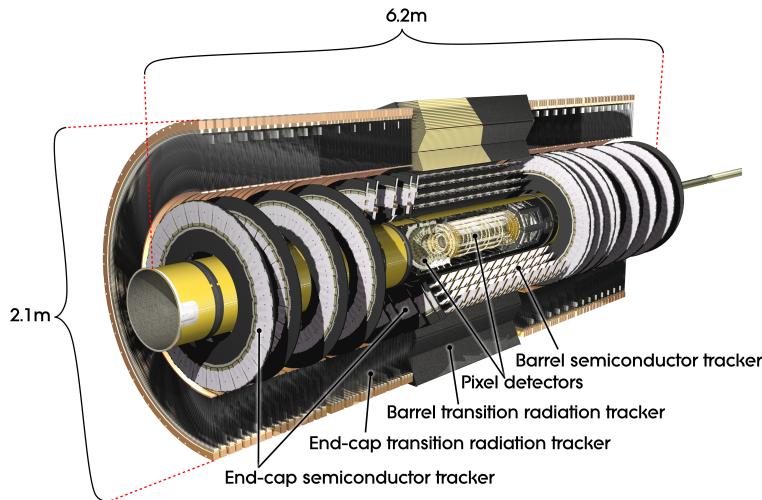


Figure 3.9: Computer generated image of the ATLAS inner detector [68].

With the exception of the TRT the tracking sub-systems are silicon based detectors. The silicon detection medium acts as a reverse bias diode. Charged particles incident on the silicon cause ionisation in the depletion layer. The products of this ionisation are electrons and holes (excess pockets of positive charge in the silicon) which produce a signal that must be handled by a read-out system. The signal is referred to as the charge collected. Application-specific integrated circuits (ASICs) are used to readout the signal performing the analogue to digital conversion. The combination of the detection medium, readout system and the printed circuit board

(PCB) on which they are joined is referred to as a module.

Pixel Detectors

There are four layers of pixel detectors that are the closest components of the ID to the beam-pipe. Pixel detectors are silicon detectors where the diodes are approximately square in shape, giving the benefit of being able to resolve hits in two directions. The design originally had three layers, each $250\ \mu\text{m}$ thick with $50\ \mu\text{m}$ by $250\ \mu\text{m}$ pixels, of oxygen doped n-type silicon crystals. During LS1 a fourth layer, closest to the beam-pipe (which was also replaced for a smaller radius version) was added. This layer is known as the insertable B-layer (IBL) [69], the motivation for its addition was to counteract degradation of original performance of the ID due to irreversible damage by radiation. As well as the inclusion of the IBL, performance degradation is mitigated by increasing the bias voltage across the pixels from 100 V (their starting voltage) to up to 600 V. Additionally the IBL being closer to the beam-pipe allows for interaction vertices to be measured more precisely. The need for better reconstruction of vertices is motivated by their role in the performance of algorithms that classify jets of activity in the detector that are initiated by B-hadrons, this is where the IBL gets its name. There are no pixel detectors in the end-caps. Each pixel is small in size which mean many can fit on one module, all requiring their own conductor for readout. The solution to this challenge is to use a complex process known as bump bonding, which is both expensive and time consuming.

Semiconductor Tracker

Next closest to the beam-pipe is the SCT whose modules have long thin strip shaped diodes. The strips provide high resolution in only a single direction. In contrast to the n-type silicon of the pixels, the strips are made from p-in-n type silicon. Each SCT module is comprised of two back to back wafers such that the orientation of the strips are offset by a small angle in order to improve coverage. Each strip is covered in a metalised layer, the strips are separated by a distance of $80\ \mu\text{m}$. A rather old

and dirty SCT module left over from the quality control testing stage of production that took place at Queen Mary University of London can be seen in figure 3.10. As can be seen in the figure the SCT modules are wire bonded to their ASICs which is cheaper and faster than the bump bonding used on the pixel modules. In order to

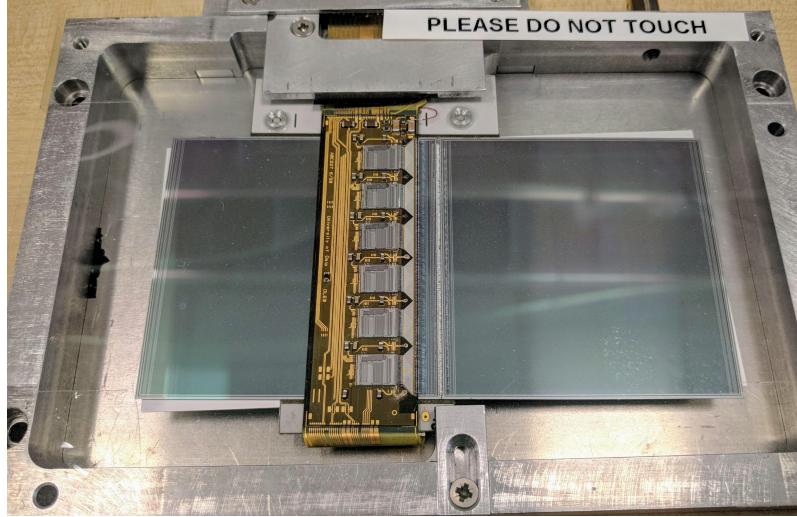


Figure 3.10: An image of an SCT long strip module mounted in a rig for testing at Queen Mary University of London.

calibrate the response of the strips a $100\text{ M}\Omega$ poly-silicon resistor is located at the end of each strip. Figure 3.11 shows an image of the snake-like structure of a poly-silicon resistor from the end of an SCT module. The modules come in two different

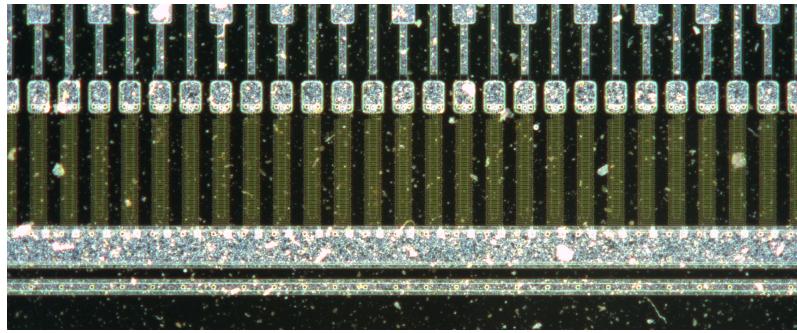


Figure 3.11: A close up image of the end of an SCT sensor in which the snake-like poly-silicon resistors are visible as a yellowish coloured structure at the end of each strip. This image was taken with a high resolution automatic area scanner commissioned by the author [70] in order to take full scans of strip sensors during the production of the ATLAS Inner Detector upgrade known as the Inner Tracker (ITk) [71, 72].

designs, short strips and long strips with the short strips forming the layer closest to the pixel detectors and the long strips on the outside. The original operating

bias voltage was 150 V but again due to radiation exposure this will raise to up to 350 V over time as necessary. There are four layers of semiconductor trackers in the barrel arranged so that sensors have a tilt with respect to a perfect coaxial cylinders of approximately 11 °. This tilt increases the amount of material that particles will travel through and is optimized to the geometry of the detector. Similarly the end-cap modules are arranged in petal like structures, with a number of different geometric designed based on the position within the end-cap.

Transition Radiation Tracker

The final layer of the ID is the TRT, the primary role of the TRT is to aid electron identification by measurement of transition radiation. The TRT is mostly made up of polyimide drift tubes with a diameter of 4 mm. The drift tubes are filled with a gas mixture whose majority constituent is xenon. These tubes operate with a voltage of -1530 V and are contained within a carbon fibre support structure. The geometric layout of the tubes is optimised individually for the barrel and end-caps.

Calorimeters

The purpose of the calorimeters is to measure the total energy of particles that pass through their volume, this is achievable only if the calorimeter stops the particle completely. A desirable side effect is that they also act as a barrier to stop particles passing through to the muon spectrometers, of course this means necessarily that muons pass through the calorimeters. There are two calorimeter systems in ATLAS the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL), they will be explained in the following sections. The calorimeters are not immersed in a significant magnetic field compared to the rest of the ATLAS as seen in the heat map of figure 3.6. The geometric layout of the calorimeter systems, as well as the location of specific components can be seen in figure 3.12, in which the ID can also be seen (greyed out). Information from the two calorimeters is used in conjunction for any particles whose decay products propagate through both volumes. Both calorimeters

are split up into cells of material that are used to determine the position of decay products in the detector.

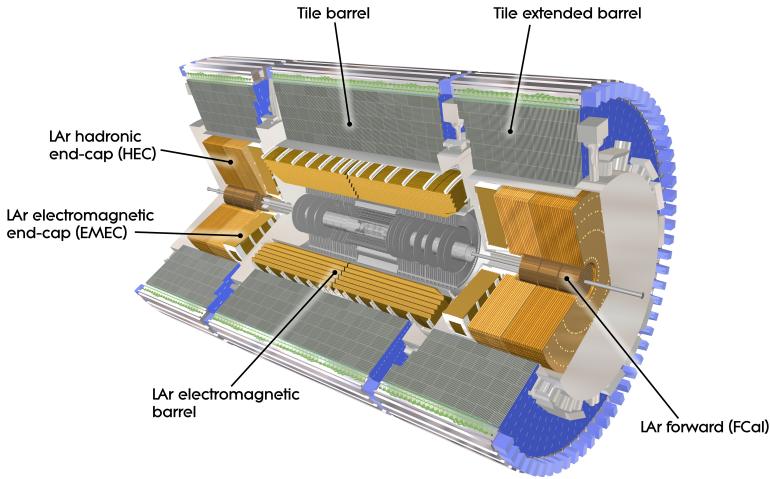


Figure 3.12: Computer Generated image of the ATLAS calorimeter [73].

High energy physics calorimeters function by measuring the shower of particles that are produced as a result of the incident particle losing its energy. Both calorimeters in ATLAS are sampling calorimeters, meaning that they are comprised of alternating layers of absorber and detection medium. The purpose of the absorber medium is to provide a material in which particle showers evolve rapidly over a short distance, though they are in general not sensitive to measuring those showers. The detection medium or active material is used to then measure those showers.

In general a good calorimeter will have the following properties. It should be hermetic meaning that there are no gaps for particles to escape through without being measured. It should have a fast response time to keep up with the rate of collisions (actually the rate of the trigger see the following section). It should be radiation tolerant in order to perform over a long period of time. It should be able to measure energy to a high resolution in order to resolve resonances over backgrounds and have a high granularity of cells in order to accurately determine the position of energy deposits (and aid particle identification). High energy resolution is greatly aided by not having anything front of the calorimeter which may change an incident particles energy before it strikes the calorimeter itself. This requirement is therefore

somewhat at odds with having a detector that has good pT resolution for electrons, positrons and photons (muons go all the way through the calorimeter anyway) due to the process by which pT is measured becoming more precise with a longer “lever-arm” distance L as seen in figure ?? and equation 3.4.

Electromagnetic Calorimeter

The ECAL is primarily concerned with measuring the energy of electrons, positrons and photons. These particles primarily interact with the electromagnetic force and so produce electromagnetic showers when they lose their energy. The ECAL resides closer to the interaction point than the HCAL and has liquid argon (LAr) as its active material. LAr is a good choice for the calorimeter which is closer to the interaction point as it is naturally resistant to damage by radiation. The absorber of the ECAL is lead, it is suitable as it has a high number of nucleons (Z) and the radiation length¹ of a given shower is inversely proportional to Z^2 . An applied electric field causes ions produced in the EM shower to drift in such a way that the signal induced is proportional to the energy deposited by the incident particle.

Hadronic Calorimeter

The HCAL end-caps use a LAr active material like the ECAL. The absorber is copper instead of lead and the dimensions are more suited to hadronic particle showers as opposed to electromagnetic. As noted before LAr is naturally radiation hard which is why it is used in the end-caps which see more radiation than the barrel regions due to almost all of the momentum of the colliding bunches of particles being in the axis of the beam-pipe. The barrel section of the hadronic calorimeter is made from scintilating tiles of active material interspersed with steel as an absorber. Scintillation is the process by which a scintillation medium produces light when energy particles travel through it. The amount of light produced is proportional to the energy of the incident particle that initiated the shower. In order to convert the

¹The radiation length X_0 is the thickness of material that reduces the energy of a particle by a factor of e the natural number.

light into a digital electrical signal photo-multiplier tubes (PMTs) are used. PMTs are able to measure very small amounts of light, even the incidence of a single photon. This is achieved by initially exploiting the photo-electric effect whereby the incident photon knocks an electron from a metallic part of the PMT. The current produced by this electron is then multiplied by a very large factor (up to 10^8 for some PMTs) by a series of dynodes (electrodes in vacuum that produce secondary emission).

Muon Spectrometers

Surrounding the calorimeters are the muon spectrometers, which form the most outer layer of the detector. Though muons are charged leptons just like electrons, their specific properties mean that dedicated muon spectrometers are required to detect them. Muons deposit far less energy per distance traveled than other particles meaning that they punch through most materials with ease. As can be seen in figure 3.13 the components of the muon spectrometers are the thin-gap chambers, cathode strip chambers, resistive plate chambers and monitor drift tubes. The barrel and end-cap toroid magnets immerse the muon spectrometers in a magnetic field which at its peak (visible in figure 3.6) has a strength of 4 T. Despite a stronger peaking magnetic field than in the solenoid observed muon tracks are often far less curved than that of their lighter cousins, the electrons. This is due to the larger mass of the muon. Muons do leave tracks in the ID and also deposit small amounts of energy in the calorimeters. Tracks in the muon spectrometers are matched up to tracks in the ID with the aid of the location of energy deposits in the calorimeters if possible. The full tracking information for muons can be used in algorithms such as overlap removal, which is used to remove muons from jets that they have been erroneously associated with by matching the muon with its ID track.

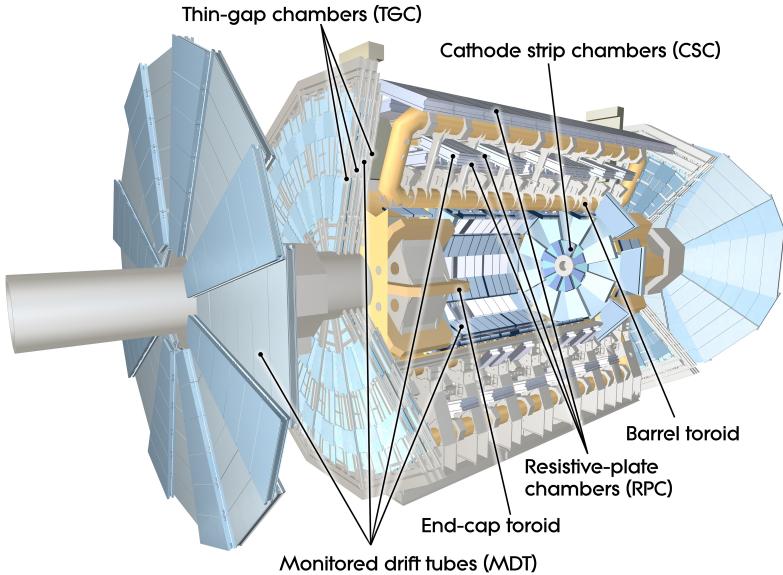


Figure 3.13: Computer generated image of the ATLAS Muons subsystem [74].

Trigger Systems

The trigger systems in ATLAS allows data to be recorded only when an event meets certain criteria. Without triggering there would be no way to decide which events to readout and which to ignore. It would be impossible to readout every interaction that occurs in the detector. The reason for this is that the geometric constraints of the detector mean that there is only a small space available for readout wires, as detection medium needs to be prioritized for sensitivity and technology limits the data rate that one can achieve through a cable of fixed area. The trigger system comes in two parts, a hardware component referred to as level one (L1) and software component referred to as the high level trigger (HLT). The L1 system is comprised of the L1 calorimeter (L1Calo) trigger which operates by searching for clusters of energy in the calorimeters and the L1 muon (L1Muon) system which coincidences in the muon systems. A third system L1 topological (L1Topo) uses regions of interest built from the L1Calo and L1Muon data which are passed to central trigger processors for selection. The various limitations of the hardware mean that these selections must be passed up to the next level of triggering, the HLT in a time window of $2.5 \mu\text{s}$. The HLT takes information from the L1 systems and uses faster versions of an offline style analysis in order to select or reject events for readout. In

order for a trigger to fire an event must pass fully all of the requirements of one of the algorithms defined by an extensive trigger menu. More information about the triggers used in the $VH(bb)$ analysis will be given in a later chapter.

Chapter 4

Machine Learning

The next chapter is somewhat of a diversion from the physics discussed so far. It will focus on techniques in machine learning which are often referred to in High Energy Physics as a Multi-variate Analysis (MVA). The reason for this diversion is that these techniques have become very widespread in the field, they are used in several of the reconstruction and selection algorithms that are used to obtain the events on which the analysis is performed. Furthermore an MVA is used to obtain the distribution that acts as the final discriminant for the analysis, and machine learning techniques are also used to model the backgrounds. Given how widespread the use of these techniques is in the analysis it makes sense to describe them before diving into the details.

The two main algorithms used are Boosted Decision Trees (BDTs) and Neural Networks (NNs) which will be described in sections 4.1 and 4.2 respectively. These algorithms are used in many places outside High Energy Physics and so rather than referring to individual pieces of data that enter into the algorithm as an event, in this chapter they will be referred to as an example. This terminology comes from the fact that in general these algorithms must be shown a large number of examples before they are suitable “trained” for their purpose, and that in general those examples could be data that represent anything. Both of these algorithms can be operated in classification or regression modes. The main difference between these modes is that classification mode provides a score for each of a given number

of classes which can be interpreted as a probability that a given example belongs to the given class, whereas regression outputs a single number per example whose interpretation depends on the problem.

4.1 Boosted Decision Trees

Decision trees have a structure as in figure 4.1.

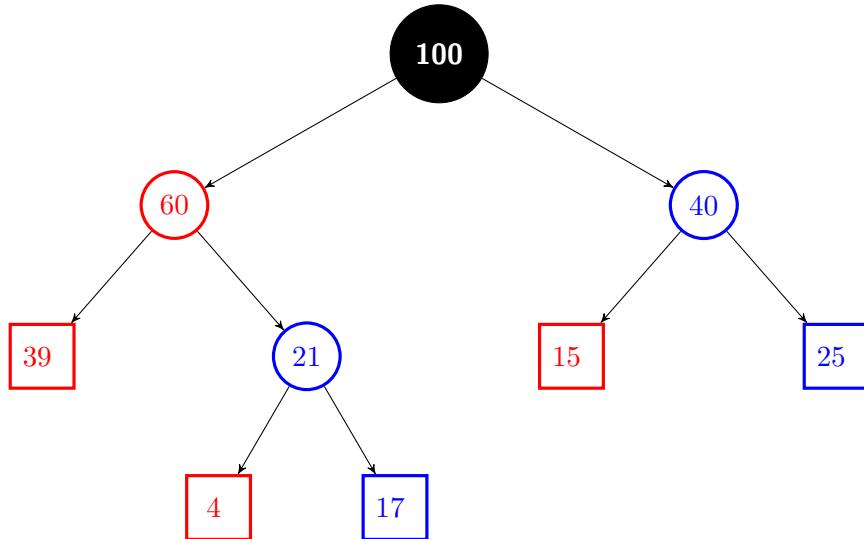


Figure 4.1: The structure of a decision tree set up for a classification problem. This diagram serves as a summary of where 100 examples end up after being passed through the tree. The number in each node corresponds to the number of examples that pass through that node.

We will start by discussing a decision tree for a classification problem. Each circular node in the tree represents a cut on one of a number of variables provided as input to the algorithm. The tree is read top to bottom with each node being followed by two edges branching left and right that represent the path taken by examples which pass or fail the cut respectively. Square nodes represent that the termination criteria have been reached and that events in these nodes have been classified according to the colour of the node. The variable chosen at each node is optimised in order to maximise a criteria related to the separation of classes. For a problem containing two classes a common separation criteria is the Gini index,

$$G = p(1 - p), \quad (4.1)$$

where p is the purity of a chosen class that one wants to maximise.

4.2 Neural Networks

Neural networks have a structure as in figure 4.4.

This section outlines some of the mathematical formalism surrounding NNs that act as classifiers. These kinds of NNs can be used to solve binary classification or multi-class problems, where the probability that each data candidate belongs to any of the classes is mutually exclusive. Many other types of NN exist, however their details will not be discussed in this work. In simple terms a NN is comprised of connected layers of nodes as in figure ???. Nodes fall into three types input, output and hidden, with layers being homogeneous with regard to node type and therefore inheriting their name. The input layer is comprised of one input node per dimension of a single entry in the dataset that one wishes to pass into the NN for classification. Each output node is a predictive unit corresponding to one of K classes and it is the goal of the intermediate hidden layers (full of hidden units) to provide a relationship between the inputs and outputs such that the predictive unit with the highest value corresponds to the correct class for the given data. The output layer should therefore be comprised of K nodes. In order to delve further into the details, a good mathematical representation of the hidden layers is required. One such representation is given by Bishop in Pattern Recognition and Machine Learning [75]. In this section the nomenclature used by Bishop will be outlined and then adapted for our specific use.

The building blocks of the NN, called activations, resemble Fisher discriminants [76] and take the form

$$a_j = \sum_{i=1}^d w_{ji}x_i + w_{j0} \quad (4.2)$$

where the w_{ji} terms are known as weights and the w_{j0} as biases. Collectively the weights and biases shall be referred to as adaptive parameters, due to the fact that

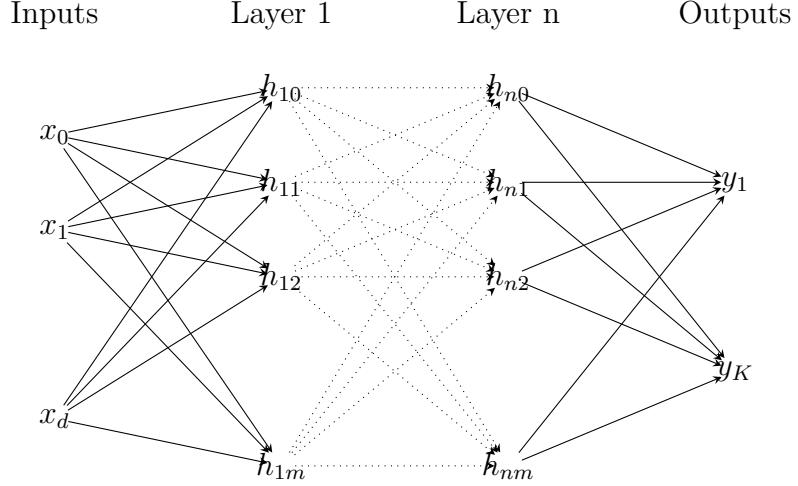


Figure 4.2: A more complex neural network containing an input layer of d nodes corresponding to data of dimensionality d , n hidden layers of m hidden units each h_{ij} (where i indexes hidden layer and j indexes a particular unit) and an output layer of K predictive units y_k .

(as shall be made clear later) they are to be adapted by a training algorithm. In order to take the activations and turn them into something closer to a perceptron [77] they must be passed through an activation function denoted \mathcal{H} ,

$$h_j = \mathcal{H}(a_j) \quad (4.3)$$

becoming what are known as hidden units. These activation functions may be non-linear and have the restriction that they must be differentiable, crucially this is different from the perceptron which uses a non-differentiable step function. Considering a network with only a single hidden layer as in figure ??, there are further steps required to get from the hidden units to the predictions y_k . The complete network function as an argument of a vector of data points \vec{x} and a matrix of adaptive parameters \vec{w} should return predictions given by

$$y_k(\vec{x}, \vec{w}) = \mathcal{O}\left(\sum_{j=1}^m w_{kj}^{(2)} \mathcal{H}\left(\sum_{i=1}^d w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right). \quad (4.4)$$

where now, as in Bishop's representation, the superscript number in brackets labels the layer to which adaptive parameters belong, not counting the input layer (as it

does not contain them). The network function (4.4) is comprised by performing the same construction on the hidden unit (4.3) as was performed on the original data point x_i in (4.2), with a special type of activation function an output function \mathcal{O} . Notably the hidden units must be summed over in the same way that data points are, however instead of summing over the dimensions of the data, in order to reproduce the network in figure ?? we sum up to m the desired number of hidden units, which is referred to as the “size” of the hidden layer. A common choice of output function for binary classification problems is the logistic sigmoid function

$$\mathcal{O}(z) = \frac{1}{1 + \exp(-z)} \quad (4.5)$$

where in each of these z merely denotes the argument of the output function. For multi-class problems where $K > 2$ a generalisation of the logistic sigmoid, the softmax function

$$\mathcal{O}(z)_k = p(k|\vec{x}) = \frac{\exp(z_k)}{\sum_{i=1}^k \exp(z_i)} \quad (4.6)$$

gives the probability of being in class k given the data \vec{x} and where i is summed all classes. The index k appears on the output function as it must be calculated for each $k \in K$ the total number of classes. It is at this stage that the true meaning of the predictive units is solidified, each should give a number between zero and one that represents the probability of data belonging to the corresponding class, and logically all predictive units should sum to one. For the sake of compactness we shall, as Bishop does, re-write (4.4) by introducing $x_0 = 1$ in order to absorb the biases into the sums, yielding

$$y_k(\vec{x}, \vec{w}) = \mathcal{O}\left(\sum_{j=0}^m w_{kj}^{(2)} \mathcal{H}\left(\sum_{i=0}^d w_{ji}^{(1)} x_i \right) \right). \quad (4.7)$$

Now it is our goal to generalise this network function to one not only of an arbitrary number of hidden layers but also such that each hidden layer can be of arbitrary size and have an arbitrary activation function. For this purpose, networks will now be described in terms of the number of hidden layers, instead of the number

of layers that contain adaptive parameters (hidden plus output) as before. We may start by writing a function for a network of two hidden layers

$$y_k(\vec{x}, \vec{w}) = \mathcal{O} \left(\sum_{j_2=0}^{m_2} w_{kj_2} \mathcal{H}_2 \left(\sum_{j_1=0}^{m_1} w_{j_2 j_1} \mathcal{H}_1 \left(\sum_{i=0}^d w_{j_1 i} x_i \right) \right) \right). \quad (4.8)$$

In order to arrive at the two layer function (4.8) the same process was used as for the original network function (4.4). Now there are two activation functions and hidden layer sizes denoted m . Clearly adding a layer to the network simply involves repeated application of this process and picking up the required number of additional parameters. A function for a network of n hidden layers may therefore be written as

$$y_k(\vec{x}, \vec{w}) = \mathcal{O} \left(\sum_{j_n=0}^{m_n} w_{kj_n} \mathcal{H}_n \left(\dots \mathcal{H}_2 \left(\sum_{j_1=0}^{m_1} w_{j_2 j_1} \mathcal{H}_1 \left(\sum_{i=0}^d w_{j_1 i} x_i \right) \right) \dots \right) \right) \quad (4.9)$$

where there are n different versions of the activation function and hidden layer size. Our new hidden units obey the following notation

$$h_{nj} = \mathcal{H}_n(a_j) \quad (4.10)$$

eliminating the need for the superscript labelling of layer. The new labelling identifies hidden layer by the left-hand index of the hidden unit or the right-hand hidden layer of the weights and biases. In equation (4.9) the hidden units are shown in their expanded form and distinguished by the fact that their j indices take on the subscript of n corresponding to the layer they belong to. Figure ?? depicts a network of n hidden layers, as described in equation (4.9) except that every hidden layer is depicted with the same size m .

There are now a great deal of parameters that we have to keep track of. It is therefore important to distinguish between adaptive parameters and the parameters which we must pick by hand, known as hyper-parameters. A relationship can be written between the number of hidden layers n , and the number of hyper-parameters,

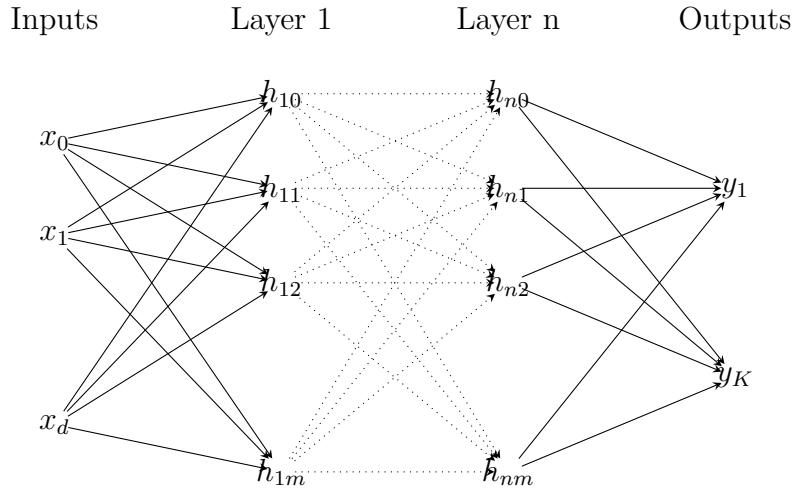


Figure 4.3: A more complex neural network containing an input layer of d nodes corresponding to data of dimensionality d , n hidden layers of m hidden units each h_{ij} (where i indexes hidden layer and j indexes a particular unit) and an output layer of K predictive units y_k .

as follows

$$\# \text{ of hyper-parameters} = 2n + 1. \quad (4.11)$$

As previously stated there are simply n lots of activation functions and n hidden layers to determine a size for. Often it is sufficient to set all activation functions to the same function, and in this work all hidden layers share the same size for simplicity. The adaptive parameters are far more numerous, and will be optimised by means of a training algorithm. In TensorFlow the variable object is choice for implementing the adaptive parameters as training algorithms in the software will update them by default. The adaptive parameters are initialised randomly from a Gaussian distribution resulting in poor predictive power to begin with. In order to improve this some figure of merit, known as a loss function, must be used in order for the training algorithm to measure quantitatively the performance of the NN. We also must provide the algorithm with a dataset to train on, complete with a set of targets or labels that, for the training set, reveal the correct classification for each entry. Due to the this requirement, methods such as these are referred to as supervised learning methods. A natural loss function one may use to describe the

error of the model given a current set of adaptive parameters is sum-of-squares error

$$E(\vec{w}) = \frac{1}{2} \sum_{n=1}^N (y(\vec{x}_n, \vec{w}) - t_n)^2 \quad (4.12)$$

where t_n are the targets for the given data entries \vec{x}_n . Minimising this function with some algorithm does work in practice, however it has been shown that, using

$$E(\vec{w}) = - \sum_{n=1}^N \left(t_n \ln(y_n) + (1 - t_n) \ln(1 - y_n) \right) \quad (4.13)$$

known as cross-entropy, is faster and more generalised [78] (further discussion on generalisation in section ??). The particular training algorithm that will be used to update parameters in this report is known as adaptive moment estimation or ADAM [79]. ADAM is a variant of the gradient descent algorithm, which is widely used and has spawned many other variants [80]. The reasons for picking between ADAM and vanilla gradient descent are given in chapter ??.

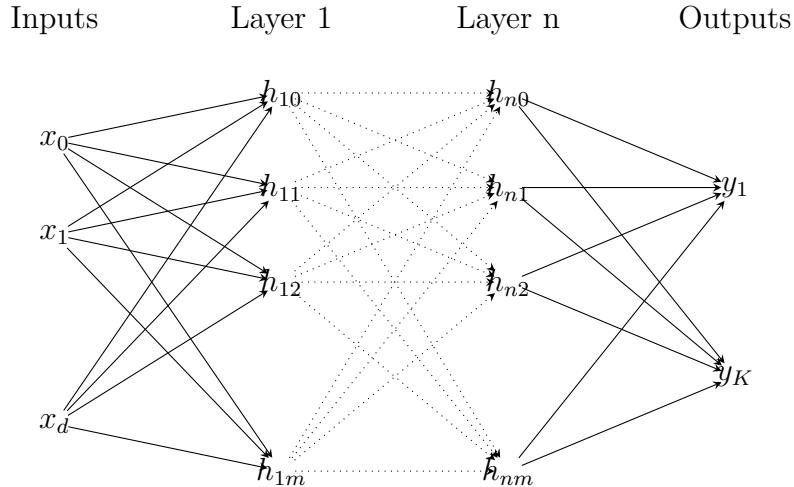


Figure 4.4: A more complex neural network containing an input layer of d nodes corresponding to data of dimensionality d , n hidden layers of m hidden units each h_{ij} (where i indexes hidden layer and j indexes a particular unit) and an output layer of K predictive units y_k .

4.3 Parametrised Neural Networks

Parametrised neural networks take extra inputs equal to the number of relevant parameters, as seen in figure 4.5.

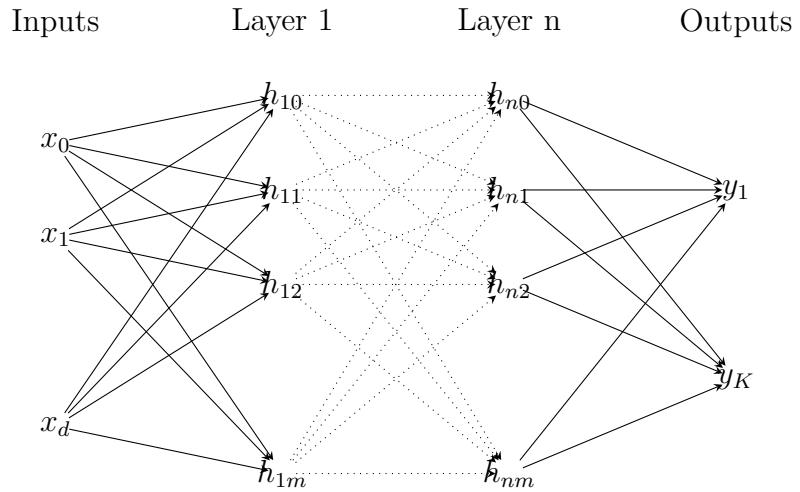


Figure 4.5: A more complex neural network containing an input layer of d nodes corresponding to data of dimensionality d , n hidden layers of m hidden units each h_{ij} (where i indexes hidden layer and j indexes a particular unit) and an output layer of K predictive units y_k .

Chapter 5

Reconstruction and Selection

At the top level events are required to have one Higgs boson candidate and one vector boson candidate. In all cases a Higgs boson candidate is comprised of two b-tagged jets. More on the jet collection and b-tagging strategy used in section 5.4. Vector boson candidates are defined by a number of different decay products defined in section ??, these decay products are triggered on, specific triggers used are discussed in section 5.3. Reconstruction of basic quantities as well as higher level algorithms such as overlap removal are handled by Athena and the CxAOD Framework, more on these in section 5.1. Finally with all quantities reconstructed events are categorised in different analysis regions, these are described in section 7.1.

5.1 Athena and the CxAOD Framework

Data recorded by the ATLAS detector is passed through the central collaboration software framework Athena before entering the analysis level data processing.

Athena is responsible for the steps shown in figure 5.1. As can be seen in the figure Athena processes both data recorded from collisions and Monte-Carlo simulated predictions. Steps up until and including reconstruction are required to transform the raw or simulated read-out of the detector into what are known as physics objects. These physics objects correspond to UV and IR safe descriptions particles and hadron showers e.g. leptons and jets. Given the initial transverse energy of the

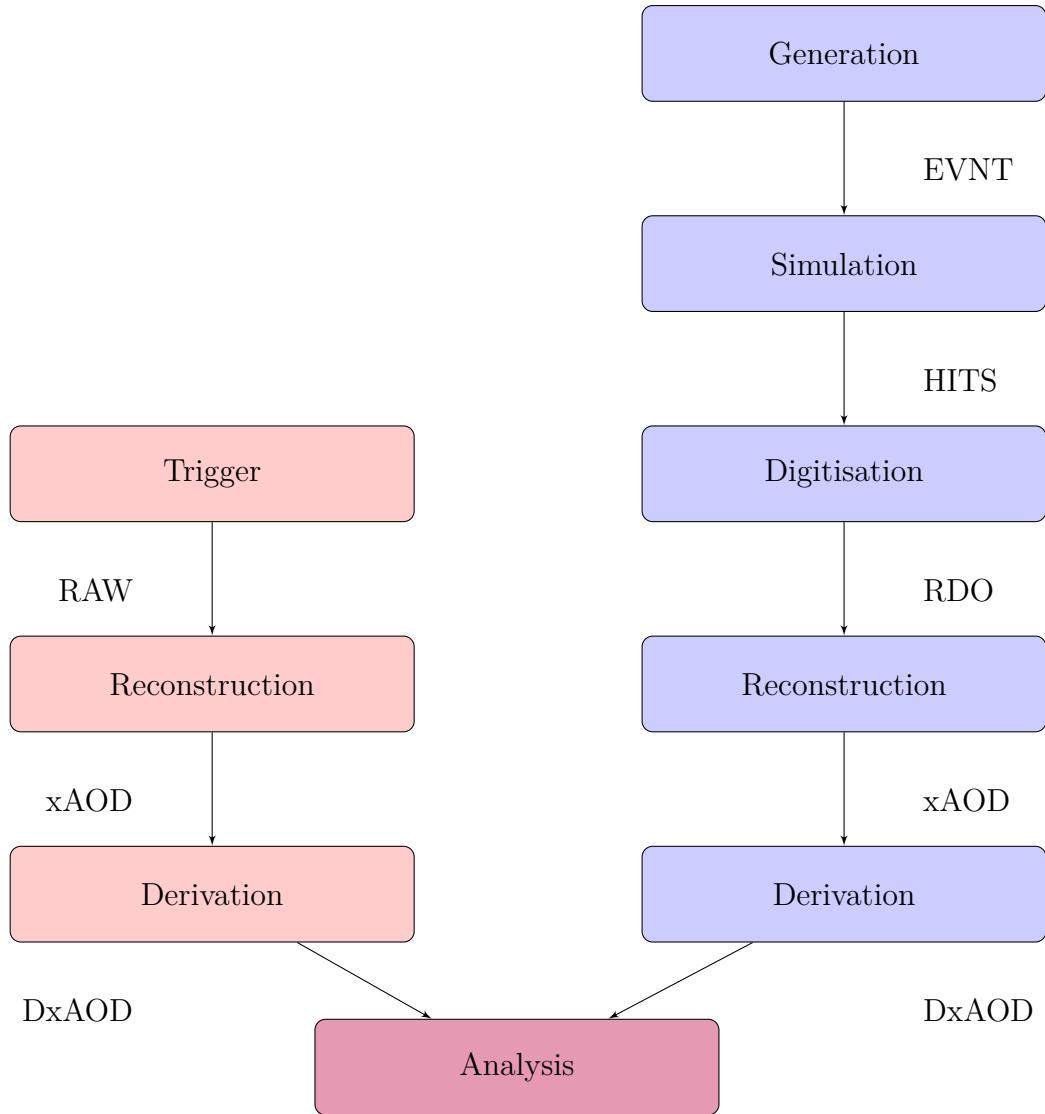


Figure 5.1: A flow chart showing the Athena data processing chain. Red nodes indicate the presence of data recorded from collisions and blue nodes indicate the presence of Monte-Carlo simulated events.

collisions (zero) any missing transverse energy (E_{miss}^T) is also reconstructed based on the sum of the transverse energy of all objects in an event, this missing energy indicates the presence of particles in the event that cannot be detected by any of ATLAS subsystems. The only particles in the Standard Model for which this is expected are neutrinos. The files containing the reconstructed physics objects adhere to the ATLAS Event Data Model (EDM) and are referred to as Analysis Object Data (xAOD). After reconstruction a part of Athena called the derivation framework is used to produce skimmed and slimmed xAODs known as Derived xAODs (DxAODs). The reduction of these files is carried out based on a loose selection

criteria, the criteria used in the VH(bb) analysis are shown in table ??.

DxAODs are the usual starting point for analysis level software, in the case of this analysis the CxAOD Framework. Of course xAODs can also be used as the starting point for analysis they are just larger. As in figure 5.2 The CxAOD Framework

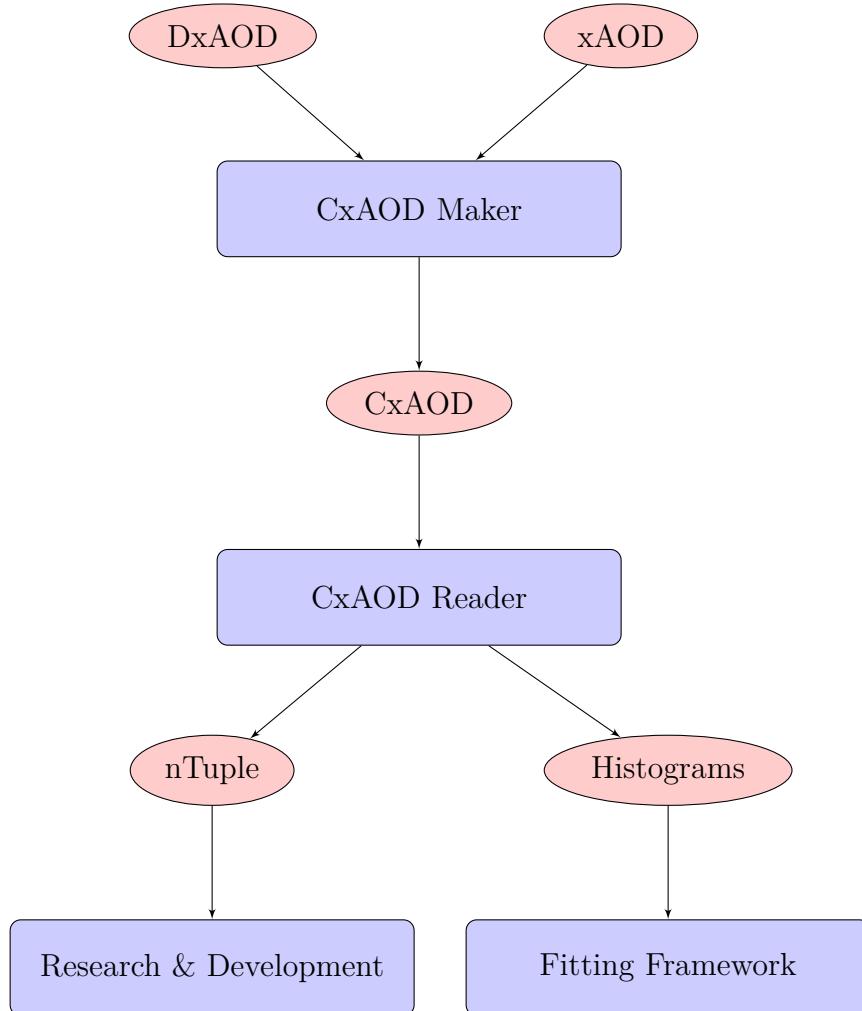


Figure 5.2: A flow chart showing the CxAOD Framework data processing chain. The red elliptical nodes indicate data formats and the blue rectangular nodes indicate software modules.

has two major components the Maker and the Reader. The job of the maker is to further slim the data by performing pre-selection cuts and also to apply calibrations which will be detailed below, the output of the Maker is called a Calibrated xAOD (CxAO). The Reader takes a CxAOD as input and performs the analysis event selection, it can output histograms or nTuples.

The Maker applies selections the selections of each of the three analysis channels, defined in section ?? . CxAODs are produced separately for each of the three

Name	p_T	$ \eta $	ID	d_0^{sig}	$ \Delta z_0 \sin $	Isolation
electrons						
VH-loose	> 7 GeV	< 2.47	LH Loose	< 5	< 0.5 mm	FCLoose
ZH-signal	> 27 GeV	< 2.47	LH Loose	< 5	< 0.5 mm	FCLoose
WH-signal	> 27 GeV	< 2.47	LH Tight	< 5	< 0.5 mm	FixedCutHighPtCaloOnly
muons						
VH-loose	> 7 GeV	< 2.7	Loose quality	< 3	< 0.5 mm	FixecdCutLoose
ZH-signal	> 27 GeV	< 2.5	Loose quality	< 3	< 0.5 mm	FixecdCutLoose
WH-signal	> 27 GeV	< 2.5	Medium quality	< 3	< 0.5 mm	FixedCutHighPtCaloOnly

Table 5.1: Definitions of the VH leptons used to define and select events for the three analysis channels where d_0^{sig} is the measured with respect to the beam line.

channels as the different background compositions and signal signatures require different optimisation. Pre-selection is performed on jets based on application of requirements on transverse momentum and pseudo-rapidity. A tool known as the Jet Vertex Tagger (JVT) is used to remove jets resulting from pileup from events.

5.2 Leptons

The channels of the VH(bb) analysis are defined by the number of observed charged leptons (e or μ) in the decay of the vector boson. There is one channel for the study of WH(bb) decays where the leptonic decay $W \rightarrow \ell\nu$ yields a single charged lepton. There are two channels for the study of ZH(bb), the zero lepton channel, $Z \rightarrow \nu\nu$, and the two lepton channel, $Z \rightarrow \ell\ell$.

Two classifications of lepton are defined in order to categorise events into the individual channels of the analysis, these are called VH-loose and VH-signal leptons, channels are kept orthogonal by requiring different numbers of both lepton categories. These classifications are defined in table 5.1. The characteristics of the fake lepton background from QCD multi-jet processes differs between the 1 and 2 lepton channels hence the reason for two different categorisations. In general to suppress this kind of background leptons are required to be isolated from other detector activity.

Electrons

As mentioned in chapter 3 electrons leave tracks in the ID and energy deposits in the ECAL. Reconstructing electrons requires clustering the energy deposits in the ECAL, this is achieved with a sliding window algorithm [?]. Clusters must then be associated to tracks in the ID, a Gaussian Sum Filter [?] is used to account for energy losses due to bremsstrahlung radiation. The energy for electron candidates must be calibrated before it can be used in order to account for things such as non-uniformity in the detector response. Calibration is achieved by using simulated cluster activity from single particles to train a BDT regression model designed to regress the measured energy in the ECAL to the simulated energy. An in-situ data driven correction is applied to normalise the response between data and simulation [?].

Reconstruction alone is not enough to find electrons, other particles may leave similar signatures in the ATLAS sub-detectors and therefore electron identification must also be performed. Identification is performed using a likelihood-based method. Variables which have power to discriminate between electrons and other particles are used in the likelihood such as shower profiles, track quality, how closely track and cluster positions match in η and ϕ , and the presence of a high-threshold TRT hit. This is one of the main benefits of the TRT. Performance of this method is well studied [?].

Muons

Finding muons in the detector requires consideration of the coverage of the different ATLAS sub-detectors, especially in general are not stopped in the detector. Muons leave charged tracks in the ID and the muon spectrometers which have coverages of $|\eta| < 2.7$ and $|\eta| < 2.5$ respectively. For the region $|\eta| > 2.5$ a stand-alone algorithm which doesn't use ID tracks can be used. All muons within the coverage of the ID require good quality ID tracks []. A combined algorithm is used in the majority of cases. For the region $|\eta| < 0.1$ two specialised algorithms SegmentTagged and

CaloTagged are used which require only muon segment and calorimeter deposits respectively. All aforementioned algorithms are used together in what is known as a unified chain [?] to reconstruct and identify muons.

Taus

As mentioned the only charged leptons that are considered in the analysis are electrons and muons, leptonically decaying taus will have electrons and muons as the only visible decay products however hadronically decaying taus must be considered differently. Decays are considered as one or three pronged based on the number of charged decay products, pions, with neutrinos and neutral pions also present. These decays are reconstructed in the calorimeters like jets with the anti- k_t algorithm with $\Delta R = 0.4$ [?] but the p_T of the tau is set to the total energy of the TopoClusters within $\Delta Ig\sigma R < 0.2$, more on TopoClusters in section 5.4. Tau candidates must have $p_T > 20 GeV$, $|\eta| < 2.5$ excluding $1.37|\eta| < 1.52$, and either exactly 1 or 3 tracks. A BDT based method for tau identification is used to reject fakes. Medium quality taus are counted for each event [?].

5.3 Triggers

As stated at the beginning of this chapter the decay products of the vector boson candidate are used to trigger the recording of events for this analysis. Therefore important triggers for the 0-lepton channel are E_T^{miss} triggers, and for the 1-lepton and 2-lepton channels the single electron, or single muon triggers. Note that it is not necessary to trigger on both charged leptons coming from the Z boson in the 2-lepton channel, the presence of one lepton allows the triggering to occur and the requirement for 2-leptons can be imposed at a later stage. Some events will be missed by not using a di-lepton trigger, however these amount to approximately only 5% of the total. The list of triggers used as they appear in the ATLAS trigger menu are shown in table 5.2, for the E_T^{miss} , electron and muon triggers respectively.

Trigger Name	Period	Threshold (GeV)	Description
E_T^{miss}			
HLT_xe70_L1XE50	2015	70 GeV	Seeded using the level L1_XE50
HLT_xe90_mht_L1XE50	2016 (A-D3)	90 GeV	(L1_XE55) LAr and Tile calorimeter triggers, calibrated at the EM
HLT_xe110_mht_L1XE50	2016 ($\geq D4$)	110 GeV	scale, with a threshold of 50(55) GeV.
HLT_xe110_pufit_L1XE55	2017	110 GeV	
HLT_xe110_pufit_xe70_L1XE50	2018	110 GeV	
electrons			
HLT_e24_lhmedium_L1EM20VH	2015	24 GeV	Seeded using L1EM20VH level 1 trigger calibrated at the EM scale with a threshold of 20 GeV, and require medium ID quality.
HLT_e60_lhmedium	2015	60 GeV	Seeded using L1EM20VH level 1 trigger calibrated at the EM scale with a threshold of 20 GeV, and require medium ID quality.
HLT_e120_lhloose	2015	120 GeV	Seeded using L1EM20VH level 1 trigger calibrated at the EM scale with a threshold of 20 GeV, and require loose ID quality.
HLT_e26_lhtight_nod0_ivarloose	2016 – 2018	26 GeV	Tight likelihood ID required, and variable loose isolation required
HLT_e60_lhmedium(_nod0)	2016 – 2018	60 GeV	Medium ID likelihood required
HLT_e140_lhloose(_nod0)	2016 – 2018	140 GeV	Loose ID likelihood required
HLT_e300_etcut	2018	300 GeV	No ID requirements.
muons			
HLT_mu20_iloose_L1MU15	2015	20 GeV	Seeded using L1MU15 level 1 trigger with a threshold of 15 GeV, and requiring loose isolation requirements.
HLT_mu50	2015 – 2018	60 GeV	No isolation requirements.
HLT_mu26_ivarmedium	2016 – 2018	26 GeV	Variable cone medium isolation requirements

Table 5.2: Triggers used during the 2015, 2016, 2017 and 2018 data collection periods, notation like A or D3 denote periods during the year.

0-lepton channel triggers

The events in the 0-lepton channel should have a $qq\nu\nu$ final state. We use the E_T^{miss} triggers listed in table 5.2 as the final state will manifest in the detector as E_T^{miss} with the presence of jets. At the stage of triggering E_T^{miss} is only calculated from energy measured in the calorimeters. As muons do not deposit much energy in the calorimeters the $W \rightarrow \mu\nu + \text{jets}$ process is used to study the trigger efficiency and derive an appropriate scale factor.

1/2-lepton channel triggers

The 1 and 2-lepton channels both contain charged leptons in the final state, their final states are specifically $qql\nu$ and $qql\ell$ respectively. As the two lepton channel is not expected to contain significant E_T^{miss} and inefficiencies in the muon trigger are mitigated by the presence of two muons in the events which wish to record the single lepton triggers listed in table 5.2 are simply used without any E_T^{miss} triggers.

In the 1-lepton channel there is significant E_T^{miss} expected due to the presence of the neutrino. Single lepton triggers are used for events with $75GeV < p_T^V < 150GeV$ where the E_T^{miss} triggers have yet to turn on fully, for events with $p_T^V > 150GeV$ in order to mitigate inefficiencies in the single muon triggers similar E_T^{miss} triggers to the 0-lepton triggers are used in conjunction with the single lepton triggers.

5.4 Jets

As mentioned in chapter 2 jets are the roughly conical structure of detector activity resulting from the ultimate hadronisation of a QCD parton. Two categories of jets are considered, signal jets and forward jets, when the number of total jets is referred to it is equal to the sum of signal and forward jets. As the Higgs candidate in every channel of the analysis is two b-tagged jets, both the reconstruction of jets and the b-tagging strategy have huge impacts on the final measurements. In this section the way jets are found and reconstructed will be introduced, b-tagging will be explained in general and then the specific tagging strategy of the analysis will be detailed.

Topological Calorimeter Cluster anti- k_t jets

The jets that are found with a given algorithm are referred to as a jet collection. The jet collection relevant to this analysis uses topological calorimeter cell clusters to reconstruct jets [?]. These cluster are then passed to the anti- k_t jet finding algorithm [81]. This algorithm takes a radius parameter which governs the size of jets, in the ATLAS coordinates a radius parameter of $R = 0.4$ is chosen.

Jet Category	Selection Requirements
Forward Jets	jet cleaning $p_T > 30 \text{ GeV}$ $2.5 \leq \eta < 4.5$
Signal Jets	jet cleaning $p_T > 20 \text{ GeV}$ $ \eta < 2.5$ JVT medium for $p_T < 120 \text{ GeV}$

Table 5.3: Jet selection requirements. Jet cleaning refers to the quality criteria interfaced through the JetCleaningTool CP tool. This tool removes jets in regions where either coherent/sporadic calorimeter noise or beam-induced background fake jets [? ?].

As mentioned in chapter 3 pileup can cause issues with reconstruction. In general there is a desire to suppress any jets which arise from pileup. The Jet Vertex Tagger (JVT) is a likelihood-based discriminant which is used to achieve this. The primary vertex location, jet p_T and the

p_T of tracks associated to a given jet, serve as inputs to the JVT which outputs a 2-D likelihood that the jet arises from pileup. The likelihood is resilient to bias arising from the jet flavour. The tool is applied only to jets in region $|\eta| < 2.5$ and $p_T > 120 \text{ GeV}$. A cut of $\text{JVT} = 0.59$, is applied to all jets in the collection, this cut has an average efficiency of 92 %. The definitions of signal and forward jets can be found in table 5.3

b-tagging

It is important to distinguish jets originating from b-quarks, which form our Higgs candidate, from c-jets and τ -jets, as well as jets originating from quarks lighter than c-quark which are categorised together as light-jets. The calculation of a discriminant which ought to separate b-jets from other jets is known as b-tagging. In order to develop such a discriminant one must use simulation in order to be able to know truly which parton initiated a jet such that the performance of the discriminant can be validated. In simulation a jet and the parton that initiated that jet are distinctly separate objects and so a set of rules must be defined in order

to decide which jet is b-jet and likewise for other types of jets. Those rules are as follows:

1. If a weakly decaying b -hadron is found within $\Delta R < R_{\max}$ of the jet axis, the jet is labeled a b -jet.
2. If a b -hadron isn't found, but a weakly decaying c -hadron is found within $\Delta R < R_{\max}$ of the jet axis, then the jet is labeled as a c -jet.
3. Otherwise, if a τ -lepton is found within $\Delta R < R_{\max}$ of the jet axis, the jet is labeled a τ -jet.
4. If any one hadron or τ -lepton matches more than one jet, the closest jet is chosen as its parent.
5. All unlabeled jets after steps 1 through 4 are labeled as light-jets.

The algorithm used to tag b-jets is the MV2c10 algorithm, this is a BDT which is trained on kinematic and structural information about each jet. It is setup to categorise between b-jets (signal) and a mixture of light-jets and c-jets (background). The events in the training sample are simulated $t\bar{t}$ events that have at least one lepton coming from a leptonically decaying W boson, and hadronically decaying Z' events. The training sample has 5 million $t\bar{t}$ events and 3 million Z' events.

The kinematic training variables that enter into the MV2c10 algorithm are simply the jet p_T and η . The structural information is more complicated, IP2D and IP3D are two algorithms based on a log-likelihood ratio discriminant of impact parameters

(see chapter 3). IP3D is defined as

$$\text{IP3D} = \sum_{i=1}^N \log \frac{p_b}{p_u} \quad (5.1)$$

where

$$p_b = P\left(\text{is b-jet} \mid \frac{d_0}{\sigma_{d_0}}, \frac{z_0 \sin \theta}{\sigma_{z_0 \sin \theta}}\right), \quad (5.2)$$

$$p_u = P\left(\text{is light-jet} \mid \frac{d_0}{\sigma_{d_0}}, \frac{z_0 \sin \theta}{\sigma_{z_0 \sin \theta}}\right), \quad (5.3)$$

and

N = the number of tracks for a given jet.

IP2D has the same definition but the probabilities p_b and p_u are conditional only on the transverse impact parameter d_0 and have no dependence on the longitudinal z_0 . The output of two algorithms designed to find secondary vertices, SV1 and JetFitter, also enter into the MV2c10 training.

A jet is defined b-tagged if its MV2c10 score exceeds a certain threshold. This threshold is defined as the cut that gives a pre-determined efficiency value for b-jets when applied to a $t\bar{t}$ sample. Calibrations are available for a number of these so-called working points, these working points are shown in table 5.4.

Name	MV2c10 cut	b-tagging efficiency (%)	c-jet rejection	light-jet rejection
FixedCutBEff_60	0.94	61.14	22	1204
FixedCutBEff_70	0.83	70.84	8	313
FixedCutBEff_77	0.64	77.52	4	113
FixedCutBEff_85	0.11	85.23	2	28

Table 5.4: b -tagging working points available in this analysis, rejection is in the inverse of efficiency.

Pseudo-continuous b-tagging

The working points defined in table 5.4 are used in a so-called pseudo-continuous mode. In this mode the MV2c10 distribution is binned with bin edges corresponding to the working points listed in the table, and as can be seen in figure ???. This figure shows the discrimination between events with different jet flavours. Events that

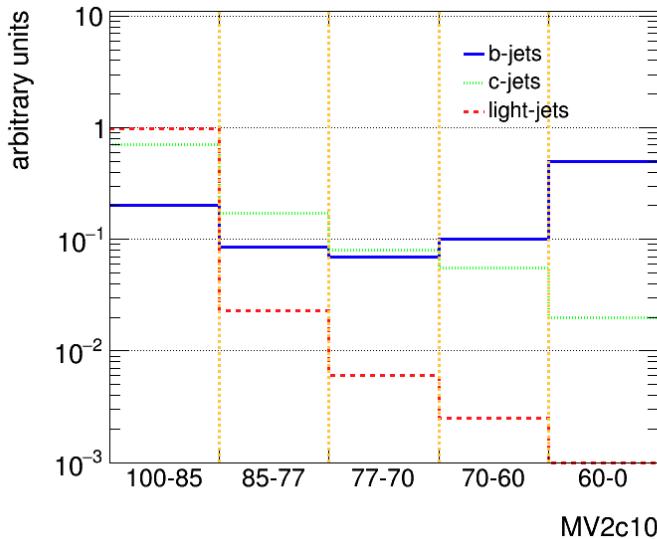


Figure 5.3: Shapes of the PCBT quantiles for b-,c-, and light-jets.

fall in the bins with ranges 70 – 60 % and 60 – 0 % have their MV2c10 score stored for use in the analysis. Events in the other bins are not tagged as b-jets.

Truth tagging

Removing all events that fail the cut associated with the 70 % working point results in loss of many events. This leaves us with low statistics samples for analysis where we would rather have high statistics samples, as many of the techniques used for categorising events have performance conditional on the available statistics. Instead of simply throwing these events away it is possible to keep all events that were simulated to have a b-hadron decay and instead weight distributions by a factor which represents the probability that an event would in fact be b-tagged and enter into the final distribution. This procedure is known as Truth-Tagging, since the information as to whether a b-hadron is truly in a simulated event is utilised, this does not use any information relating to the true nature of the data as this is unknowable. In order to draw easy comparisons, tagging directly with the MV2c10 algorithm is referred to as direct tagging.

Given that our events contain more than one jet in all cases it is required to develop a way of generating Truth-Tag weights for events based on all the jets in the event.

This weight is calculated as the product of the b-tagging efficiency for each b-tagged jet, multiplied by the complement of the b-tagging efficiency for each non b-tagged jet. In case the number of jets in a given event (m) exceeds the number of required tagged jets in the analysis ($n = 2$) all possible combinations of jets which satisfy the analysis selection are considered.

Given an event with m jets and n tagged jets required, the possible combinations of tagged and non-tagged jets are $\binom{m}{n}$. For a given tagged configuration, referred to as the i^{th} combination and denoted as $\binom{m}{n}_i$, the total number of remaining configurations is $\overline{\binom{m}{n}} = \binom{m}{n} - \binom{m}{n}_i$. The efficiency and inefficiency products are

$$\varepsilon\left(\binom{m}{n}_i, x, f\right) = \prod_{j \in n} \epsilon_x^f(j) \quad (5.4)$$

$$\varepsilon_{in}\left(\overline{\binom{m}{n}}_i, x, f\right) = \prod_{j \in m-n} (1 - \epsilon_x^f(j)) \quad (5.5)$$

with $\epsilon_x^f(j)$ the tagging efficiency of jet j of flavour f at an efficiency working point x , and where $j \in m - n$ refers to the pool of non tagged jets. The total event weight is

$$w_{TT}(x) = \sum_i^{\binom{m}{n}} \varepsilon\left(\binom{m}{n}_i, x\right) \cdot \varepsilon_{in}\left(\overline{\binom{m}{n}}_i, x\right), \quad (5.6)$$

with a probability to choose a specific combination equal to

$$P_i(x) = \frac{\varepsilon\left(\binom{m}{n}_i, x\right) \cdot \varepsilon_{in}\left(\overline{\binom{m}{n}}_i, x\right)}{w_{TT}}. \quad (5.7)$$

A single truth tagged combination can be selected based on this probability, and the event is then scaled by the factor w_{TT} . In practice due to differences between the cumulative and pseudo-continuous b-tagging efficiency distributions a modified version of Truth-Tagging must be applied in the analysis, the modifications are described in appendix 9.1.

Hybrid tagging

There exists non-closure between direct-tagged and truth-tagged events which means that truth tagging cannot directly be applied to events in the analysis. This is because direct tagging is the only strategy which can be used on data and so we can only assume that direct-tagged distributions in simulation will describe the shape and normalisation of distributions in the data. This problem is solved by implementing so-called hybrid tagging. The hybrid tagging strategy involves the following steps:

1. Divide the jets of each event in two groups, depending on the truth tag flavour:
a group with only true b-jets and the other with non b-jets.
2. All b-jets in the first group are direct tagged.
3. The remaining group of c- and light-jets is truth tagged imposing a number of tagged jets proportional to the difference between the number of required jets in the signal region and the number of true b-jets in the first group passing the b-tag requirement.

Distributions of $t\bar{t}$ and $W + \text{jets}$ are shown in figures 5.4 and 5.5. Which show direct, truth and hybrid-tagged distributions. It can be seen that the closure of hybrid-tagged events is much better than that of truth-tagged events with respect to that of the direct-tagged events. Hybrid tagging is therefore chosen as the analysis strategy.

5.5 Missing Transverse Momentum

This section describes how E_T^{miss} is calculated. The calculation is based on the assumption that partons entering into the hard scatter interaction have zero transverse momentum when they collide. This is not true due them having Fermi temperature, but the transverse momentum induced by this phenomenon is minimal and therefore is ignored. Given this assumption E_T^{miss} is calculated as the negative vector sum

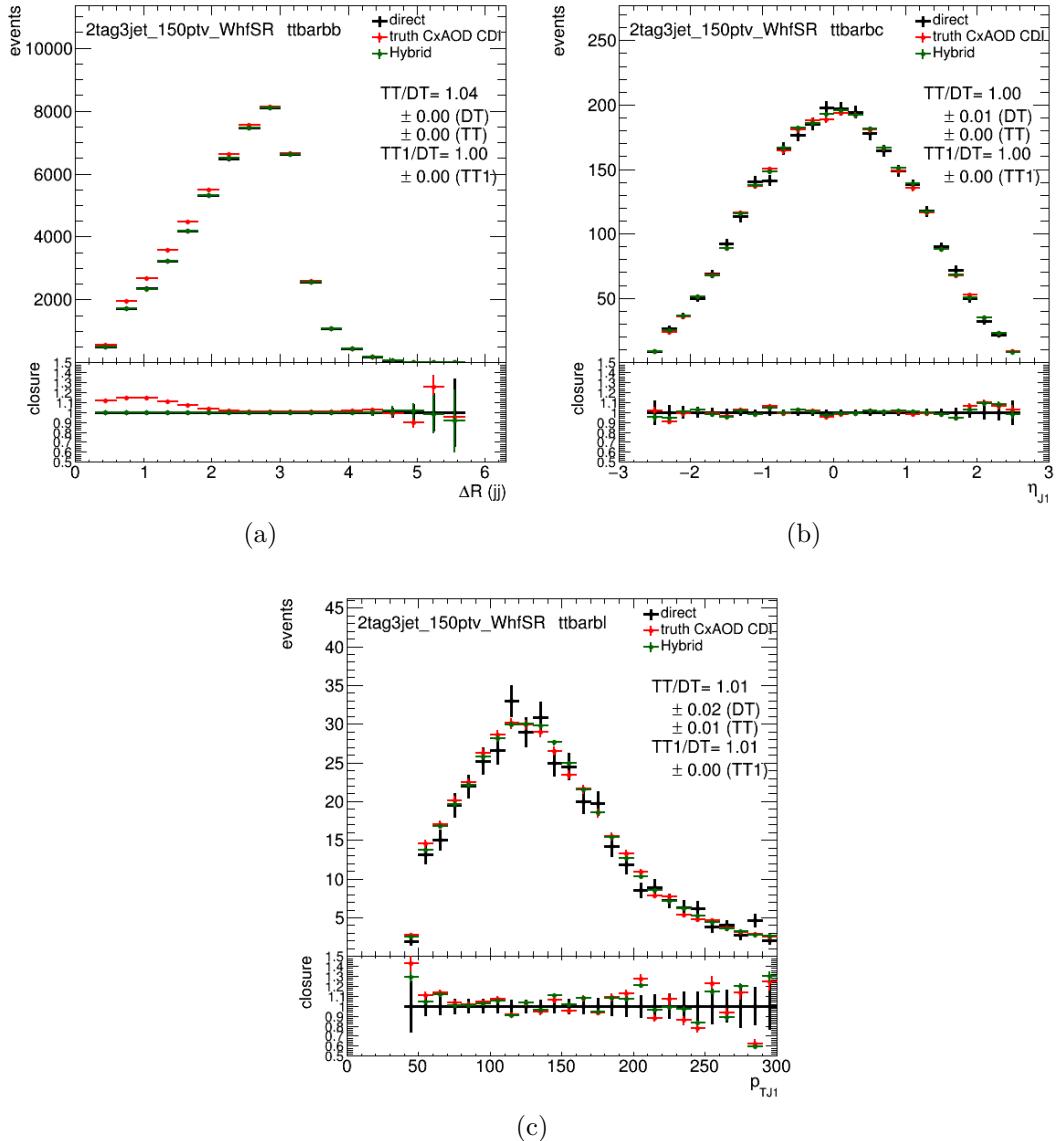


Figure 5.4: (a) $\Delta R(\text{jet}, \text{jet})$ distribution in $t\bar{t}(bb)$ events, (b) η distribution in $t\bar{t}(bc)$ events and (c) p_T distribution of the leading jet in $t\bar{t}(bl)$ events, using (black) direct tagging, (red) truth tagging and (green) hybrid tagging with the new customized maps.

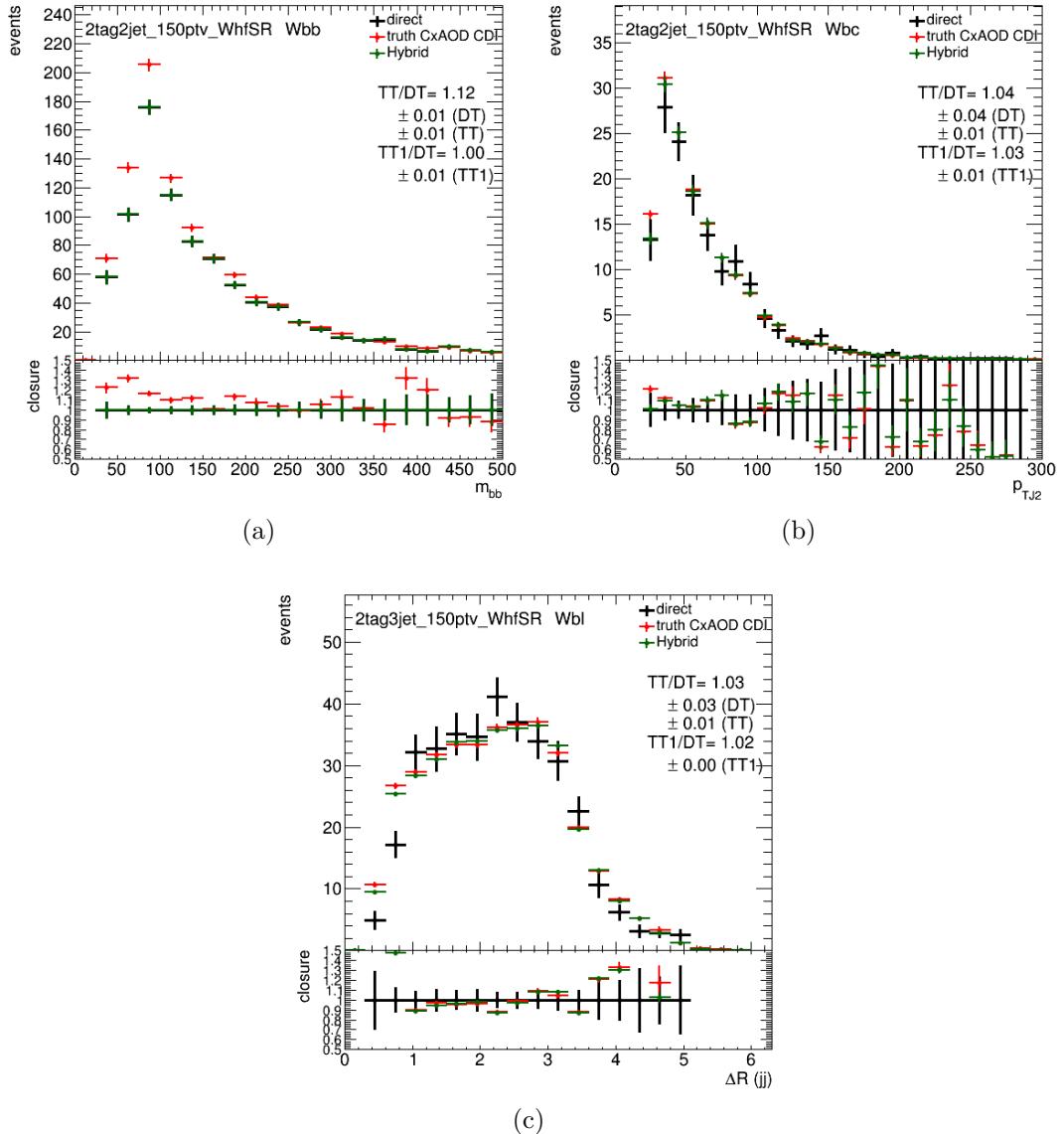


Figure 5.5: (a) m_{bb} distribution in $W+bb$ events, (b) p_T of the sub-leading jet for $W+bc$ events and (c) $\Delta R(jet, jet)$ distribution for $W+bl$ events using (black) direct tagging, (red) truth tagging and (green) hybrid tagging with the new customized maps.

of the p_T of photons, electrons, muons, taus, and jets. There is also an additional soft term that enters into the sum which is made up of all good quality tracks that aren't associated with any of the aforementioned objects. These tracks must be associated with the primary vertex and therefore are robust against pileup. A second formulation of E_T^{miss} called $E_{T,Trk}^{miss}$ is formulated just using inner detector tracks, it is even more robust to pileup but cannot account for neutral particles which do no leave tracks in the inner detector. One more variable known as E_T^{miss} -significance is also calculated as

$$E_T^{miss}\text{-significance} = \frac{E_T^{miss}}{\sqrt{\sum p_T^e + \sum p_T^\mu + \sum p_T^{jet}}}, \quad (5.8)$$

where the denominator is clearly proportional to the E_T^{miss} resolution and therefore making cuts on this variable simulates making harder cuts on E_T^{miss} with better resolution and looser cuts on E_T^{miss} with poorer resolution.

5.6 Overlap Removal

Given all of the reconstruction methods described so far for different physics objects there has yet been any mention of what happens if overlapping detector activity could be reconstructed into more than one physics object. The treatment of overlap depends on which physics objects are involved. Below is summary of how different pairs of overlapping objects which are relevant to the analysis are handled, in the summary a combined muon is one which has been reconstructed with the combined muon reconstruction algorithm, that is to say it has tracks in the ID and energy deposits in the ECAL.

- **tau-electron:** If $\Delta R(\tau, e) < 0.2$, the τ lepton is removed.
- **tau-muon:** If $\Delta R(\tau, \mu) < 0.2$, the τ lepton is removed, unless the τ lepton has $p_T > 50$ GeV and the muon is not a combined muon, then the τ lepton is not removed.
- **electron-muon:** If a combined muon shares an ID track with an electron, the

electron is removed. If a calo-tagged muon shares an ID track with an electron, the muon is removed.

- **electron-jet:** If $\Delta R(\text{jet}, e) < 0.2$ the jet is removed. For any remaining jets, if $\Delta R(\text{jet}, e) < 0.4$, the electron is removed.
- **muon-jet** If $\Delta R(\text{jet}, \mu) < 0.2$ or the muon ID track is ghost associated to the jet, then the jet is removed if the following also holds. The jet has less than three associated tracks with $p_T > 500$ MeV or the p_T ratio of the muon and jet is larger than 0.5 and the ratio of the muon p_T to the sum of p_T of tracks with $p_T > 500$ MeV associated to the jet is larger than 0.7. For any remaining jets, if $\Delta R(\text{jet}, \mu) < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^\mu)$, the muon is removed from the jet.
- **tau-jet:** If $\Delta R(\tau, \text{jet}) < 0.2$, the jet is removed.

5.7 Final Selection

The final analysis selection varies between lepton channels. There are however some common selections to all three channels as the Higgs candidate does not differ between channels, after all this is what we aim to measure. All events must have at least two signal jets. These signal jets must be b-tagged, the b-tagging algorithm used is the MV2c10 algorithm with the 70% efficiency working point used, events with between 0 or 1 b-tags are considered for study and are not used in the final analysis, events with ≥ 3 b-tags are rejected entirely. The leading b-tagged jet may have $p_T > 45$ GeV. A key variable in the analysis, the di-jet mass, m_{jj} is reconstructed from the two leading jets, in the case where both jets are b-tagged $m_{bb} \equiv m_{jj}$ and represent the mass of the Higgs candidate.

Reconstruction of the momentum of b-tagged jets can be enhanced with better resolution with respect to other jets. This is achieved with so-called muon-in-jet and p_T -reco corrections detailed in ???. Corrections are applied after events pass the full analysis selection, but before other stages of the analysis. These corrections are not used in the calculation of any E_T^{miss} related variables.

0-lepton channel selection

The vector boson candidate in the 0-lepton channel is a Z boson decaying to two neutrinos. In order to select events for this candidate a large amount of E_T^{miss} is required (> 150 GeV) and there must be exactly 0 VH-loose leptons in the event. Whilst events with lower E_T^{miss} may come from the physical process we desire to measure the E_T^{miss} trigger thresholds require setting the cut at it's value, the efficiency is 90% for events with $E_T^{miss} = 150$ GeV and efficiency plateaus at $E_T^{miss} \approx 180$ GeV.

There exists a non-trivial dependence of the trigger efficiency on activity of jets in detector. This arises due to the fact that the calculation of E_T^{miss} is simply the total transverse energy of the event minus the transverse energy of all objects in the event. These effects are hard to model in certain phase spaces and so a requirement is put on S_T such that it must be greater than 120 GeV for events with two total jets, and greater than 150 GeV for events with three total jets, where S_T is defined as,

$$S_T = \sum_i p_T^i, \text{ for } i \text{ jets in the event.} \quad (5.9)$$

Due to the lack of charged leptons in the $Z \rightarrow \nu\nu$ decay there is nothing to trigger on to suppress multi-jet background processes arising from QCD. This background is also enhanced due to limitations of the calorimeter performance. A set of so-called anti-QCD cuts are applied to deal with this multi-jet background, they are as follows:

- $|\Delta\Phi(E_T^{miss}, E_{t,trk}^{miss})| < 90^\circ$
- $|\Delta\Phi(j_1, j_2)| < 140^\circ$
- $|\Delta\Phi(E_T^{miss}, h)| > 120^\circ$
- $\min(|\Delta\Phi(E_T^{miss}, \text{pre-sel. jets})|) > 20^\circ \text{ for 2 jets, } > 30^\circ \text{ for 3 jets.}$

Where $E_{t,trk}^{miss}$ is defined as the missing transverse momentum calculated from the negative vector sum of the transverse momenta of tracks reconstructed in the inner

detector and identified as originating from the primary vertex. The leading and sub-leading jets are denoted j_1 and j_2 . These cuts reduce the multi-jet background to approximately 1% of the total background in this channel.

1-lepton channel selection

The vector boson candidate in the 1-lepton channel is a W boson decaying to one charged lepton and one neutrino. In order to select for events of this signature exactly one WH-signal lepton is required. At low p_T there is increased contribution from multi-jet processes, therefore the extra requirements of $p_T^W > 150$ GeV and $E_T^{miss} > 30$ GeV. The latter cut is only applied in the electron channel.

2-lepton channel selection

The vector boson candidate in the 1-lepton channel is a W boson decaying to two charged leptons. Exactly two VH-loose leptons of the same lepton flavour are required, additionally one of the leptons must pass the ZH-signal requirements. For events with two muons, it is required that the muons are of opposite charge. Electron reconstruction suffers from a higher rate of charge misidentification and so this requirement is not applied to events with two electrons. The di-lepton invariant mass is confined to be around the Z boson mass peak, events require $81 < m_{ll} < 101$ GeV. These cuts reduce multi-jet backgrounds to negligible levels.

The presence of two visible leptons in this channel allows for the enhancement to b-tagged jets momentum resolution to be further improved. By inspection of the decay in the channel it is clear that the momentum of the Higgs and vector boson candidates ought to be balanced. The momentum resolution of the two charged leptons forming the Z boson candidate is higher than that of the b-tagged jets forming the Higgs candidate (even after corrections). The b-tagged jets momentum can therefore be corrected with a kinematic fit. The kinematic fit improves resolution by 20-30% compared with the muon-in-jet corrected quantities. This correction is

only applied to events with 2 or 3 total jets as the presence of more jets results in smearing of the effect over those additional jets.

Common Selections

Jets	≥ 2 signal jets
b -jets	2 b -tagged signal jets
Leading b -tagged-jet p_T	> 45 GeV

0 Lepton

Trigger	lowest un-prescaled E_T^{miss} triggers
Leptons	0 VH-loose lepton
E_T^{miss}	> 150 GeV
S_T	> 120 (2 jets), > 150 GeV (3 jets)
$ \min \Delta\phi(E_T^{miss}, \text{jet}) $	$> 20^\circ$ (2 jets), $> 30^\circ$ (3 jets)
$ \Delta\phi(E_T^{miss}, h) $	$> 120^\circ$
$ \Delta\phi(j_1, j_2) $	$< 140^\circ$
$ \Delta\phi(E_T^{miss}, E_{T,\text{trk}}^{miss}) $	$< 90^\circ$
p_T^V regions	[150, 250] GeV, [250, ∞] GeV

1 Lepton

Trigger	e channel: un-prescaled single electron Tables 6 and 7 of Ref. [?] μ channel: see 0-lepton triggers
Leptons	1 WH-signal lepton > 1 VH-loose lepton veto
E_T^{miss}	> 30 GeV (e channel)
p_T^V regions	[150, 250] GeV, [250, ∞] GeV

2 Lepton

Trigger	un-prescaled single lepton Tables 6 and 7 of Ref. [?]
Leptons	2 VH-loose leptons (≥ 1 ZH-signal lepton) Same flavor, opposite-charge for $\mu\mu$
m_{ll}	$81 < m_{ll} < 101$ GeV
p_T^V regions	[75, 150], [150, 250], [250, ∞] GeV

Table 5.5: Summary of the signal event selection in the 0-, 1- and 2-lepton analyses.

Chapter 6

Fit Models

6.1 Profile Likelihood Fits

6.2 VH(b,b) multi-variate discriminant fit

6.3 Di-jet mass fit

6.4 VZ Cross-check fit

Chapter 7

Background Modelling

7.1 Categorisation into Analysis Regions

Top $e\mu$ control region

$\Delta R(b, b)$ Control Regions

7.2 Novel Modelling Techniques

Multi-dimensional re-weighting

Inferring missing information with parametrised neural networks

7.3 $Z + \text{jets}$ modelling

7.4 Modelling of other backgrounds

Variation	Values		
Sherpa 2.2.1			
Factorisation scale (μ_F)	$2\mu_F$	$0.5\mu_F$	
Renormalisation scale (μ_R)	$2\mu_R$	$0.5\mu_R$	
PDF Variation	MMHT2014nnlo68cl	CT14nnlo	
Sherpa 2.1			
Re-summation scale (μ_S)	$2\mu_S$	$0.5\mu_S$	
CKKW Merging scale	15 GeV	30 GeV	

Table 7.1: A summary of the Sherpa 2.2.1 and Sherpa 2.1 internal variations that are used to model V + jets processes.

Chapter 8

Results

**8.1 VH(b,b) multi-variate discriminant fit
results**

8.2 Di-jet mass fit results

8.3 VZ Cross-check results

Chapter 9

Conclusion

9.1 Future studies

q

Bibliography

- [1] Georges Aad et al. (ATLAS Collaboration). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.
- [2] Serguei Chatrchyan et al. (CMS Collaboration). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.
- [3] Will J. Percival, Shaun Cole, Daniel J. Eisenstein, Robert C. Nichol, John A. Peacock, Adrian C. Pope, and Alexander S. Szalay. Measuring the baryon acoustic oscillation scale using the Sloan Digital Sky Survey and 2df Galaxy Redshift Survey. *Monthly Notices of the Royal Astronomical Society*, 381(3):1053–1066, 2007. doi: 10.1111/j.1365-2966.2007.12268.x. URL <http://dx.doi.org/10.1111/j.1365-2966.2007.12268.x>.
- [4] Fabio Iocco, Gianpiero Mangano, Gennaro Miele, Ofelia Pisanti, and Pasquale D. Serpico. Primordial Nucleosynthesis: from precision cosmology to fundamental physics. *Phys. Rept.*, 472:1–76, 2009. doi: 10.1016/j.physrep.2009.02.002.
- [5] M. Kowalski et al. Improved Cosmological Constraints from New, Old and Combined Supernova Datasets. *Astrophys. J.*, 686:749–778, 2008. doi: 10.1086/589937.
- [6] Richard Massey et al. Dark matter maps reveal cosmic scaffolding. *Nature*, 445:286, 2007. doi: 10.1038/nature05497.

- [7] Douglas Clowe, Marusa Bradac, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *Astrophys. J.*, 648:L109–L113, 2006. doi: 10.1086/508162.
- [8] J. Anthony Tyson, Greg P. Kochanski, and Ian P. Dell’Antonio. Detailed mass map of CL0024+1654 from strong lensing. *Astrophys. J.*, 498:L107, 1998. doi: 10.1086/311314.
- [9] Lars Bergstrom. Dark Matter Candidates. *New J. Phys.*, 11:105006, 2009. doi: 10.1088/1367-2630/11/10/105006.
- [10] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016. doi: 10.1088/1674-1137/40/10/100001.
- [11] F. Zwicky. Republication of: The redshift of extragalactic nebulae. *General Relativity and Gravitation*, 41:207–224, January 2009. doi: 10.1007/s10714-008-0707-4.
- [12] Lars Bergström. Nonbaryonic dark matter: Observational evidence and detection methods. *Rept. Prog. Phys.*, 63:793, 2000. doi: 10.1088/0034-4885/63/5/2r3.
- [13] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.*, 405:279–390, 2005. doi: 10.1016/j.physrep.2004.08.031.
- [14] Roel Aaij et al. Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\mu^-\bar{\nu}_\mu)$. *Phys. Rev. Lett.*, 115(11):111803, 2015. doi: 10.1103/PhysRevLett.115.159901, 10.1103/PhysRevLett.115.111803. [Erratum: *Phys. Rev. Lett.* 115,no.15,159901(2015)].
- [15] J. P. Lees et al. Evidence for an excess of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ decays. *Phys. Rev. Lett.*, 109:101802, 2012. doi: 10.1103/PhysRevLett.109.101802.

- [16] Thomas Blum, Achim Denig, Ivan Logashenko, Eduardo de Rafael, B. Lee Roberts, Thomas Teubner, and Graziano Venanzoni. The Muon ($g-2$) Theory Value: Present and Future. 2013.
- [17] Carl E. Carlson. The Proton Radius Puzzle. *Prog. Part. Nucl. Phys.*, 82:59–77, 2015. doi: 10.1016/j.ppnp.2015.01.002.
- [18] Bernat Capdevila, Andreas Crivellin, Sébastien Descotes-Genon, Joaquim Matias, and Javier Virto. Patterns of New Physics in $b \rightarrow s\ell^+\ell^-$ transitions in the light of recent data. *JHEP*, 01:093, 2018. doi: 10.1007/JHEP01(2018)093.
- [19] John D. Hobbs, Mark S. Neubauer, and Scott Willenbrock. Tests of the standard electroweak model at the energy frontier. *Rev. Mod. Phys.*, 84:1477–1526, Oct 2012. doi: 10.1103/RevModPhys.84.1477. URL <https://link.aps.org/doi/10.1103/RevModPhys.84.1477>.
- [20] L. Nodulman. Experimental tests of the standard model. *NATO Sci. Ser. C*, 534:147–193, 1999. doi: 10.1007/978-94-011-4689-0_4.
- [21] Morad Aaboud et al. Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector. *Phys. Lett.*, B786:59–86, 2018. doi: 10.1016/j.physletb.2018.09.013.
- [22] Francis Halzen, Alan D. Martin, and Nilotpal Mitra. Quarks and leptons: An introductory course in modern particle physics. *American Journal of Physics*, 53(3):287–287, 1985. doi: 10.1119/1.14146. URL <https://doi.org/10.1119/1.14146>.
- [23] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. doi: 10.1017/CBO9781139525367.
- [24] A. Zannoni. On the Quantization of the Monoatomic Ideal Gas. *eprint arXiv:cond-mat/9912229*, December 1999.
- [25] On the theory of quantum mechanics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 112(762):

- 661–677, 1926. ISSN 0950-1207. doi: 10.1098/rspa.1926.0133. URL <http://rspa.royalsocietypublishing.org/content/112/762/661>.
- [26] A. Borrelli. *Compendium of Quantum Physics*. 2009.
- [27] Emmy Noether. Invariant variation problems. *Transport Theory and Statistical Physics*, 1(3):186–207, Jan 1971. ISSN 1532-2424. doi: 10.1080/00411457108231446. URL <http://dx.doi.org/10.1080/00411457108231446>.
- [28] Albert Einstein. On the electrodynamics of moving bodies. *Annalen Phys.*, 17: 891–921, 1905. doi: 10.1002/andp.200590006. [Annalen Phys.14,194(2005)].
- [29] V. Alan Kostelecký and Neil Russell. Data tables for lorentz and *cpt* violation. *Rev. Mod. Phys.*, 83:11–31, Mar 2011. doi: 10.1103/RevModPhys.83.11. URL <https://link.aps.org/doi/10.1103/RevModPhys.83.11>.
- [30] David Mattingly. Modern tests of lorentz invariance. *Living Reviews in Relativity*, 8(1):5, Sep 2005. ISSN 1433-8351. doi: 10.12942/lrr-2005-5. URL <https://doi.org/10.12942/lrr-2005-5>.
- [31] S Liberati. Tests of lorentz invariance: a 2013 update. *Classical and Quantum Gravity*, 30(13):133001, jun 2013. doi: 10.1088/0264-9381/30/13/133001. URL <https://doi.org/10.1088%2F0264-9381%2F30%2F13%2F133001>.
- [32] Clifford M. Will Mark P. Haugan. Modern tests of special relativity. *Physics Today*, 40(5):69, 1987. doi: 10.1063/1.881074. URL <https://doi.org/10.1063/1.881074>.
- [33] Floyd W. Stecker. Limiting superluminal electron and neutrino velocities using the 2010 crab nebula flare and the icecube pev neutrino events. *Astroparticle Physics*, 56:16 – 18, 2014. ISSN 0927-6505. doi: <https://doi.org/10.1016/j.astropartphys.2014.02.007>. URL <http://www.sciencedirect.com/science/article/pii/S092765051400019X>.

- [34] R. Abbasi et al. Search for a lorentz-violating sidereal signal with atmospheric neutrinos in icecube. *Phys. Rev. D*, 82:112003, Dec 2010. doi: 10.1103/PhysRevD.82.112003. URL <https://link.aps.org/doi/10.1103/PhysRevD.82.112003>.
- [35] Enrico Borriello, Sovan Chakraborty, Alessandro Mirizzi, and Pasquale Dario Serpico. Stringent constraint on neutrino lorentz invariance violation from the two icecube pev neutrinos. *Phys. Rev. D*, 87:116009, Jun 2013. doi: 10.1103/PhysRevD.87.116009. URL <https://link.aps.org/doi/10.1103/PhysRevD.87.116009>.
- [36] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.*, 28:1049–1070, Dec 1926. doi: 10.1103/PhysRev.28.1049. URL <https://link.aps.org/doi/10.1103/PhysRev.28.1049>.
- [37] Sheldon L. Glashow. The renormalizability of vector meson interactions. *Nucl. Phys.*, 10:107–117, 1959. doi: 10.1016/0029-5582(59)90196-8.
- [38] Abdus Salam and John Clive Ward. Weak and electromagnetic interactions. *Nuovo Cim.*, 11:568–577, 1959. doi: 10.1007/BF02726525.
- [39] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967. doi: 10.1103/PhysRevLett.19.1264.
- [40] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957. doi: 10.1103/PhysRev.105.1413. URL <https://link.aps.org/doi/10.1103/PhysRev.105.1413>.
- [41] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $s^{**}(1/2) = 540\text{-GeV}$. *Phys. Lett.*, B122:103–116, 1983. doi: 10.1016/0370-2693(83)91177-2. [,611(1983)].
- [42] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p

- Collider. *Phys. Lett.*, B122:476–485, 1983. doi: 10.1016/0370-2693(83)91605-2. [,7.45(1983)].
- [43] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider. *Phys. Lett.*, B126:398–410, 1983. doi: 10.1016/0370-2693(83)90188-0. [,7.55(1983)].
- [44] P. Bagnaia et al. Evidence for $Z_0 \rightarrow e^+ e^-$ at the CERN anti-p p Col- linder. *Phys. Lett.*, B129:130–140, 1983. doi: 10.1016/0370-2693(83)90744-X. [,7.69(1983)].
- [45] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964. doi: 10.1103/PhysRevLett. 13.321. URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [46] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. doi: 10.1103/PhysRevLett.13.508. URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [47] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964. doi: 10.1103/PhysRevLett.13.585. URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.585>.
- [48] P. W. Anderson. Plasmons, gauge invariance, and mass. *Phys. Rev.*, 130: 439–442, Apr 1963. doi: 10.1103/PhysRev.130.439. URL <https://link.aps.org/doi/10.1103/PhysRev.130.439>.
- [49] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016. doi: 10.23731/CYRM-2017-002.
- [50] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. URL <https://cds.cern.ch/record/782076>.

- [51] *LEP design report.* CERN, Geneva, 1984. URL <https://cds.cern.ch/record/102083>. Copies shelved as reports in LEP, PS and SPS libraries.
- [52] CERN. ATLAS: Letter of intent for a general purpose p p experiment at the large hadron collider at CERN. 1992.
- [53] M. Della Negra et al. CMS: letter of intent by the CMS Collaboration for a general purpose detector at LHC. Technical Report CERN-LHCC-92-003. LHCC-I-1, CERN, Geneva, 1992. URL <https://cds.cern.ch/record/290808>. Open presentation to the LHCC 5 November 1992, M. Della Negra/CERN, CMS Spokesman.
- [54] H. Dijkstra, Hans Jürgen Hilke, Tatsuya Nakada, and Thomas Ypsilantis. LHCb Letter of Intent, LHCb Collaboration. 1995.
- [55] Letter of Intent for A Large Ion Collider Experiment [ALICE]. Technical Report CERN-LHCC-93-016. LHCC-I-4, CERN, Geneva, 1993. URL <https://cds.cern.ch/record/290825>.
- [56] G Giacomelli, A A Faust, and James L Pinfold. A search for highly ionizing particles and slow exotic decays at the LHC using the MOEDAL detectors: letter of intent. Technical Report CERN-LHCC-98-05. LHCC-I-19, CERN, Geneva, Feb 1998. URL <https://cds.cern.ch/record/347906>.
- [57] W Kienzle, , et al. *Total cross section: elastic scattering and diffraction dissociation at the LHC.* Letter of Intent. CERN, Geneva, 1997. URL <http://cds.cern.ch/record/335255>.
- [58] O Adriani. LHCf Letter of Intent for a p-Pb run. A precise study of forward physics in $\sqrt{s_{NN}} = 4.4$ TeV proton-Lead ion collisions with LHCf at the LHC. Technical Report CERN-LHCC-2011-015. LHCC-I-021, CERN, Geneva, Dec 2011. URL <https://cds.cern.ch/record/1404163>.
- [59] Christiane Lefèvre. The CERN accelerator complex. Complexe des accélérateurs du CERN. Dec 2008. URL <https://cds.cern.ch/record/1260465>.

- [60] Design Report Tevatron 1 project. Technical Report FERMILAB-DESIGN-1984-01, 1984. URL <https://cds.cern.ch/record/1478620>.
- [61] I. Bejar Alonso and L. Rossi. HiLumi LHC Technical Design Report. 2015.
- [62] Georges Aad et al. (ATLAS Collaboration). The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, 2008. URL <http://stacks.iop.org/1748-0221/3/i=08/a=S08003>.
- [63] David Hanser. *Architecture of France*. Greenwood Press, Westport, Conn, 2006. ISBN 978-0-313-31902-0.
- [64] Joao Pequenao. Computer generated image of the whole ATLAS detector. Mar 2008. URL <https://cds.cern.ch/record/1095924>.
- [65] A. Airapetian et al. ATLAS: Detector and physics performance technical design report. Volume 1. 1999.
- [66] A. Airapetian et al. ATLAS: Detector and physics performance technical design report. Volume 2. 1999.
- [67] H.H.J. ten Kate. The atlas superconducting magnet system at the large hadron collider. *Physica C: Superconductivity*, 468(15):2137 – 2142, 2008. ISSN 0921-4534. doi: <https://doi.org/10.1016/j.physc.2008.05.146>. URL <http://www.sciencedirect.com/science/article/pii/S0921453408004541>. Proceedings of the 20th International Symposium on Superconductivity (ISS 2007).
- [68] Joao Pequenao. Computer generated image of the ATLAS inner detector. Mar 2008. URL <https://cds.cern.ch/record/1095926>.
- [69] M Capeans, G Darbo, K Einsweiller, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne, and R Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010. URL <https://cds.cern.ch/record/1291633>.

- [70] Thomas Paul Charman, Bart Hommels, and Adrian John Bevan. Authorship Qualification Task: ATLAS ITk Sensor Scanner Documentation (QMUL). Technical Report ATL-COM-ITK-2019-018, CERN, Geneva, Jan 2019. URL <https://cds.cern.ch/record/2674404>.
- [71] Ben Brueers. The design and layout of the Phase-II upgrade of the Inner tracker of the ATLAS experiment. Jul 2019. URL <https://cds.cern.ch/record/2681301>.
- [72] John Stakely Keller. The ATLAS ITk Strip Detector System for the Phase-II LHC Upgrade. Feb 2019. URL <https://cds.cern.ch/record/2663172>.
- [73] Joao Pequenao. Computer Generated image of the ATLAS calorimeter. Mar 2008. URL <https://cds.cern.ch/record/1095927>.
- [74] Joao Pequenao. Computer generated image of the ATLAS Muons subsystem. Mar 2008. URL <https://cds.cern.ch/record/1095929>.
- [75] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.
- [76] R. A. FISHER. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936. ISSN 2050-1439. doi: 10.1111/j.1469-1809.1936.tb02137.x. URL <http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x>.
- [77] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, pages 65–386, 1958.
- [78] John Platt Patrice Y. Simard, Dave Steinkraus. Best practices for convolutional neural networks applied to visual document analysis. Institute of Electrical and Electronics Engineers, Inc., August 2003. URL <https://www.microsoft.com/en-us/research/publication/best-practices-for-convolutional-neural-networks-applied-to-visual-document-analysis/>

[practices-for-convolutional-neural-networks-applied-to-visual-document-analysis/](https://arxiv.org/abs/1704.04367).

- [79] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL <http://arxiv.org/abs/1412.6980>.
- [80] Sebastian Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016. URL <http://arxiv.org/abs/1609.04747>.
- [81] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

Truth tagging for pseudo-continuous b-tagging

Truth tagging has been extended to work in the pseudo-continuous tagging strategy mentioned in section 5.4. As was explained, truth tagging of the tagged and non-tagged jets are decided randomly a priori, and the assigned weight reflects the probability for the jets to be above the tagging threshold. However, by definition, in pseudo-continuous working point there is no a priori tagging threshold, which needs to be specified as an additional parameter.

Furthermore, there is a fundamental difference between the definition of flavour-tagging efficiency in the cumulative and pseudo-continuous working points. In the first case the efficiency defines the probability for a jet to have a value *above* the b-tagging requirement, while in the second case the efficiency quantifies the probability for a jet to fall *in one specific bin* of the pseudo-continuous distribution.

Thus, an additional step is needed to convert this definition of efficiency in probability to *pass* a cumulative tagging requirement. In the pseudo-continuous working point the efficiency maps are provided as 3-D maps in $(p_T, \eta, \text{weight})$, where weight is the b-tagging information divided in 5 bins. The probability for a jet to fall in a particular b-tagging bin (i_{op}) is

$$\text{Eff}_{\text{bin}}^{i_{op}} = \text{MCeff}_{\text{bin}}^{i_{op}} \cdot \text{SF}_{\text{bin}}^{i_{op}}, \quad (9.1)$$

with $\text{MCeff}_{\text{bin}}^{i_{op}}$ the MC efficiency of that particular bin, taken from the efficiency maps, and $\text{SF}_{\text{bin}}^{i_{op}}$ the flavour tagging scale factor associated to the bin. In order to recover the same definition of efficiency as the cumulative working point, the bins from 1 to i_{op} needs to be summed together as

$$\text{Eff}_{\text{cut}}^{i_{op}} = \sum_1^{i_{op}} \text{Eff}_{\text{bin}}^{i_{op}}. \quad (9.2)$$

For example, the probability for a jet to pass the 70% operating point is given by the sum of the efficiencies in the 70 – 60 % and 60 – 0 % bins, each of them corrected for the scale factor of that particular bin taken from the pseudo-continuous working point.

The scale factors are defined in a different way with respect to the cumulative working point. In practice, they are a non-trivial extension of the efficiency and inefficiency scale factors used for the regular calibrations, modified so that the tag weight fractions (both in data and in MC) sum up to unity for each kinematic bin.