Measurement of the associated production of a top quark pair and a Higgs boson ($t\bar{t}H$) with boosted topologies

Albert György Borbély

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

February 2023

Abstract

This thesis presents three studies focusing on boosted topologies that utilise machine learning techniques for boosted $H \to b\bar{b}$ reconstruction using the ATLAS detector. The measurement of the $t\bar{t}H$ cross-section is a direct way of accessing the Higgs top Yukawa coupling (y_t). Firstly, an all-hadronic feasibility study is shown, aimed at assessing boosted topologies in the all-hadronic $t\bar{t}H$ decay channel. It was found to have low statistical significance, with considerable efforts and data driven techniques required to reduce the QCD-multijet background. Secondly, the boosted contribution to the recent $t\bar{t}H$, $H \to b\bar{b}$ measurement using the full Run-2 ATLAS data set, $139fb^{-1}$ at $\sqrt{s} = 13$ TeV, is analysed. There is a considerable contribution from the boosted region to this result, particularly to the differential cross-section measurement of the Simplified Template Cross-Section (STXS) bins $[300, 450)$ and $[450, \infty)$ GeV. The result of the inclusive profile-likelihood fit is $\mu = 0.35^{+0.36}_{-0.34}$, which corresponds to $\sigma = 1.0(2.7)$ observed(expected) significance compared to the background-only hypothesis. Thirdly work on retraining the boosted $H \to b\bar{b}$ reconstruction deep neural network (DNN) is shown for the Run-2 Legacy re-analysis. The bespoke DNN trained for the analysis showed some improvements over the previous round due to the updated analysis algorithms. It also outperformed the general purpose $H \to b\bar{b}$ Xbb tagger. The final motivation for use of the bespoke DNN is that it allows the choice of boosted jet collection (RC-jets vs LR-jets). RC-jets re-cluster "small" ($\Delta R = 0.4$) jets with $\Delta R = 1.0$ while LR-jets directly cluster the calorimeter clusters with $\Delta R = 1.0$, both using the anti-k_t algorithm. The RC-jets jets are found to be advantageous. This is due to the ease of propagating systematics for combining with resolved regions and the good modelling observed using samples made with the Atlfast-2 detector simulation.

Contents

List of Tables

List of Figures

Declaration

The research studies presented in this thesis are the results of my own work within the Experimental Particle Physics group in the School of Physics and Astronomy at the University of Glasgow. It has not been submitted for another qualification at this or any other university.

The research was conducted as part of the ATLAS collaboration at the LHC and makes use of the tools and expertise developed by the collaboration. The list below outlines which parts of the thesis contain my own original work:

- Chapter 4 talks about my qualification task, required for ATLAS authorship, where I worked on chip metrology intended for use in the production life cycle of the new ITk strip production.
- Chapter 8 contains my analysis work and can be broken down into three main $t\bar{t}H$ studies, these are:
  - Section 8.1 is an all hadronic feasibility study where I investigated the feasibility of a purely boosted $t\bar{t}H$ semi-leptonic analysis.
  - Section 8.2 is a summary of the Lepton + Jets Full Run-2 Analysis [1] with a focus on the boosted semi-leptonic region, to which I directly contributed.
  - Section 8.3 describes the beginnings of the Lepton + Jets Legacy Analysis, which is a re-analysis of the Full Run-2 ATLAS data set, where my work on retraining the Higgs classifier for the boosted semi-leptonic region is detailed. The Legacy Analysis is currently in review seeking unblinding approval.

Chapter 1

Introduction

Creating, detecting and then identifying particles is the primary goal in particle physics. This allows us to probe the Standard Model (SM) which models the physical world in terms of particles, including three of the fundamental forces of nature (electromagnetic, weak and strong nuclear forces), which are the force propagators between particles. More specifically the SM is a quantum field theory (QFT) where each fundamental particle has an associated quantum field with it, of which a particle is in-fact an excitation, alluding to the wave-particle duality of quantum mechanics. Unification of quantum mechanics and special-relativity requires a quantum field theory in order to describe phenomena such as the creation and annihilation of particles and insuring causality.

In order to test the SM volumes of space with an extremely high energy density are required. This is to replicate certain conditions after the big bang when more exotic matter existed. In our current every day world the energy density is relatively low and so only the lightest particles are stable enough to exist day to day. However, as particles are the excitations of fields an observation of a specific particle, that matches the predicted properties, is a direct proof of the existence of that field. The particle properties that can be predicted from theory are: spin, electric charge, colour charge and to some degree "generation". The masses of the particles are not, although can be constrained to some degree.

In order to properly probe the SM a reliable way of producing particles is required, which is where colliders come into play. Because the SM is inherently probabilistic all matter may spontaneously be created and then annihilate as long as it doesn't violate the uncertainty principle. In order to observe a particle it needs to be created in a detector and live long enough to become "real" (on-shell). For this to happen there must be enough energy in a region of space to satisfy said particles rest mass, which in our low energy density universe will not produce any heavy particles. To get around this colliders are used which smash particles together at extremely high energies to create a localised region of space with a high energy density. This then allows, due to Einstein's energy mass

relation ($e = mc^2$) , to produce particles of higher rest mass. Which in turn provides evidence of the associated quantum field. The probability of creating a given particle (cross-section) depends on a number of things, including the particles mass (m) and the centre of mass energy ($\sqrt{s}$) of the collision, and is a measurable quantity.

The Large Hadron Collider (LHC) is a 27 km ring which accelerates protons (hadrons) to collide at 4 primary interaction points: ATLAS, LHCb, ALICE and CMS. It accelerates protons in bunches to speeds of $99.999999\%c$ and produces collisions every $25ns$. In order to produce the heavier particles in the SM a centre of mass energy higher than the rest mass of said particles must be achieved $\sqrt{s} > m$. Due to the circular shape the limiting factor, for the LHC, in the achievable centre of mass energy is the magnet technology and its achievable field strength. This limits the inward acceleration of the particles required for a given radius. To precisely know the centre of mass energy of the collision, the momentum fraction of the colliding constituents must be known. In the case of point like particles this is simple (100%) and is why initially electrons and positrons were used by LEP (Large electron positron collider). However, due to the electrons light mass, the maximum centre of mass energy that the LEP could achieve was 209 GeV. Switching to proton-proton collisions allows to achieve higher energies, 13 TeV in Run-2, but carries an additional uncertainty as the proton is not a point like particle and is made up of constituent partons. Because of this the precise momentum fraction of each parton colliding is not known. DIS (deep inelastic scattering) experiments are used to probe the structure of the proton.

The ATLAS detector, along with CMS, is one of the two general purpose detectors at the LHC. Their primary goals were to discover new particles, which at the time included the Higgs boson, as well as to be able to probe particularly high energy sectors of the SM, specifically the Higgs sector. This was because lighter masses of the Higgs had been ruled out from previous colliders such as LEP and the Tevatron.

One aspect of the Higgs sector that is of particular interest is the coupling strength to the top field (top Yukawa coupling y_t). It is particularly hard to access due to the top quark having a high mass. Typically the Yukawa coupling strengths can be accessed through the Higgs decay and their branching ratios, which is not a possibility for the top quark. Instead it can accessed by measuring the cross-section of an associated production process, where both the top-quark and Higgs boson are produced. The best process for this with the highest cross-section is the associated production of a pair of top quarks and the Higgs boson ($t\bar{t}H$).

The rest of this document is laid out as follows: Chapter 2 gives a brief overview of the SM; Chapter 3 gives an overview of the LHC and the ATLAS detector; Chapter 4 gives details on Strip Module Metrology for the ITk Inner Detector upgrade for ATLAS targeting the High Luminosity Run-4; Chapter 5 gives an overview of how theoretical predictions are made via Monte-Carlo Simulation; Chapter 6 explains the collection of

statistical methods and machine learning techniques utilised in the analysis; Chapter 7 describes the algorithms utilised by ATLAS to reconstruct the non-analysis specific objects; Chapter 8 describes the analyses targeting the $t\bar{t}H$ process with the $H \to b\bar{b}$ decay mode; and Chapter 9 gives a conclusion. The analysis chapter, Chapter 8, is split into three separate studies: an all-hadronic feasibility study, an overview of the recently published Lepton + Jets $t\bar{t}H$, $H \to b\bar{b}$ paper using the full Run-2 ATLAS data set and work for the re-analysis Legacy paper which uses the same Run-2 ATLAS data set but incorporates updated analysis techniques.

Chapter 2

The Standard Model

The SM is the theory that describes the fundamental particles and the forces between them. It is a relativistic quantum field theory (QFT) and makes use of the axioms of both special relativity and quantum mechanics. The elementary particles in the SM can be seen in Figure 2.1 and consist of three generations of fermions (quarks and leptons), four force mediating gauge bosons (g, γ, Z^0 and $W^{\pm}$) and an additional scalar boson (Higgs).

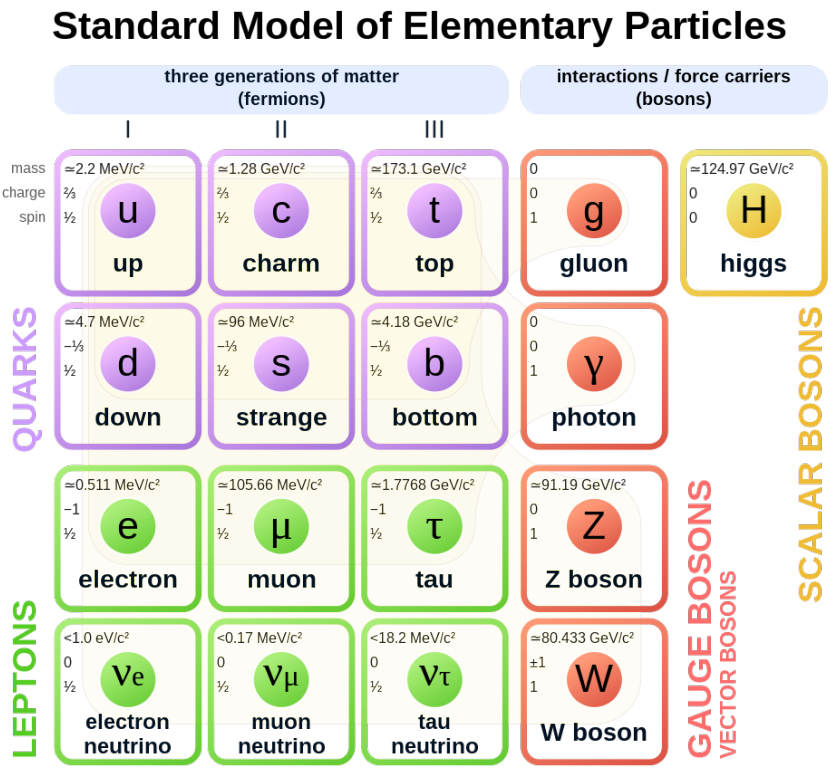


Figure 2.1: The Standard Model of elementary particles [2].

Their interactions are described by the SM Lagrangian:

$$\mathcal{L}_{SM} = \mathcal{L}_{Gauge} + \mathcal{L}_{Fermions} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}$$

which can be decomposed into various sectors. There are a total of 19 free parameters in the SM that are not predicted theoretically and need to be experimentally measured via pp collisions at the LHC.[1] The SM is a gauge theory meaning it is invariant under non-local (gauge) transformations. Its gauge symmetry group is:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y.$$

Noether's theorem [3] broadly states that for every symmetry there is a corresponding conserved quantity. For the SM gauge symmetry shown above the conserved quantities are the colour charge C SU(3), weak isospin L SU(2) and weak hypercharge Y U(1).

2.1 Electroweak Unification

The electroweak force (EW) is the unified description of the electromagnetic (EM) and weak nuclear (W) forces. It was developed Glashow, Weinberg and Salam [4–7] and is known as the Weinberg-Salam theory. Above the unification energy ~ 246 GeV (temperature 10^{15}K) the two forces merge into one. The gauge bosons are $\boldsymbol{W} = (W_1, W_2, W_3)$ for SU(2) and B for U(1), which are initially massless. The generators for the groups are weak isospin (L) and hypercharge (Y) which, due to the Higgs mechanism and spontaneous symmetry breaking, gives rise to the electric charge which can be expressed as a linear combination of the two $Q = \frac{1}{2}Y - L_z$, where L_z is the third component of weak isospin. The helicity of a particle is determined by the axis of spin being in either the same (right handed) or opposite (left handed) direction to its momentum. It was experimentally observed that weak charged currents (mediated by the $W^{\pm}$ boson) only couple to left-handed fermions thus only left handed fermions carry isospin. The fields of the physical bosons: $W^{\pm}$, Z^0 and γ expressed in terms of $\boldsymbol{W}$ and B are:

$$W^{\pm} = \frac{1}{\sqrt{2}}(W_1 \mp iW_2),$$

$$Z^0 = -B\sin(\theta_w) + W_3\cos(\theta_w),$$

$$\gamma = B\cos(\theta_w) + W_3\sin(\theta_w),$$

where θ_w is the weak mixing angle and can be defined in terms of the mass ratio of the $W^{\pm}$ and Z boson $\cos\theta_w = \frac{m_w}{m_z}$. The $W^{\pm}$ and Z^0 bosons acquire mass from the Higgs mechanism and as a result of spontaneous symmetry breaking.

[1] This can be extended with the addition of the parameters related to neutrinos to 26.

2.2 The Higgs Mechanism

$$\mathcal{L}_{Higgs} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi)$$

The Higgs Mechanism [8] is responsible for generating the masses of particles. The massive vector gauge bosons gain mass due to the spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y$. A complex scalar field $\boldsymbol{\phi} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ is introduced with a potential of the form $V(\phi^\dagger\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2$, where μ^2 and λ are free parameters. When $\mu^2 < 0$ and $\lambda > 0$ then the minimum value is $\nu = -\mu^2/2\lambda$ and is known as the vacuum expectation value of the field. Due to the Mexican hat like shape of the potential (see Figure 2.2) this minimum value ν is not at the origin ($\phi = 0$), which is in fact an unstable local maximum, but has a circle (with radius $r = \sqrt{-\mu^2/2\lambda}$) of possible values that ϕ can take. Due to the field arbitrarily "selecting" a value of ϕ to reach the minimum its symmetry becomes spontaneously broken. Because of this mass terms appear for the $W^\pm$ and Z^0 vector gauge bosons along with another massive scalar boson. It is this complex scalar field that is the Higgs field and it is this extra massive scalar boson, that appears as an excitation of this field, that is the Higgs boson. The Higgs boson has a mass $m_H = \sqrt{-2\nu^2}$ and the observed vacuum expectation value of the Higgs field is $\nu \approx 246$ GeV. It is through this symmetry breaking and the interactions with the fermions that the fermions, also gain mass. The mechanism that describes this is the Higgs fields interaction with the $\mathcal{L}_{Yukawa}$ piece of the SM Lagrangian. The fermion masses are proportional to their Yukawa couplings $y_f = \sqrt{2}m_f/\nu$, which are their couplings to the Higgs field.

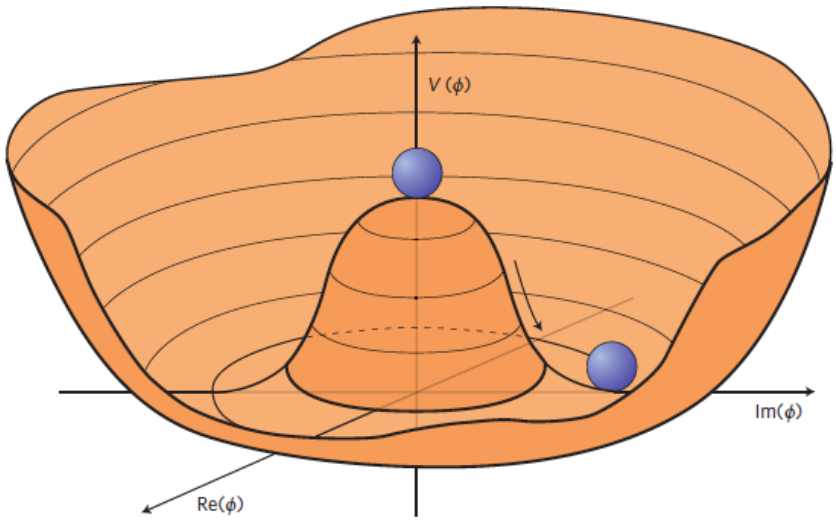


Figure 2.2: The Mexican hat shaped Higgs potential [9].

2.2.1 The Higgs boson

The Higgs boson is the mediating particle of the scalar Higgs field. It was discovered in 2012 by both ATLAS [10] and CMS [11] with its mass being measured to be $m_H = 125.25 \pm 0.17$ GeV [12], and is consistent with the limits placed on it by other experiments i.e. Tevatron. The four primary Higgs production modes in pp collisions at the LHC are gluon-gluon fusion (ggF), vector boson fusion (VBF), associated production with a vector boson (VH) and associated production with a top quark pair ($t\bar{t}H$), their Feynman diagrams can be seen in Figure 2.4. The Higgs boson production cross-sections as a function of the centre of mass energy $\sqrt{s}$ can be seen in Figure 2.3. It primarily decays to two b quarks ($\sim 58\%$) but was detected through its $H \to \gamma\gamma$ ($\sim 0.23\%$), $H \to ZZ^* \to 4l$ ($\sim 2.6\%$) and $H \to WW^* \to e\nu\mu\nu$ ($\sim 21\%$) [2] decays due to them providing a particularly clean signal. The branching fractions can be seen in Figure 2.5 which shows the probability for Higgs bosons decays through various processes at different mass values.

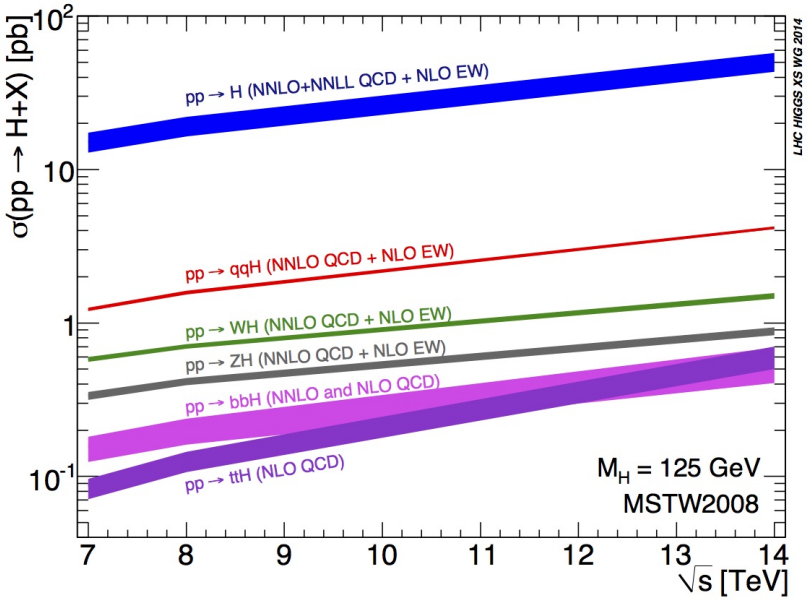


Figure 2.3: The SM Higgs boson production cross-sections as a function of centre of mass energy $\sqrt{s}$ [13].

[2] WW^* and ZZ^* decays are only possible if one of the vector bosons is produced off shell and so are suppressed

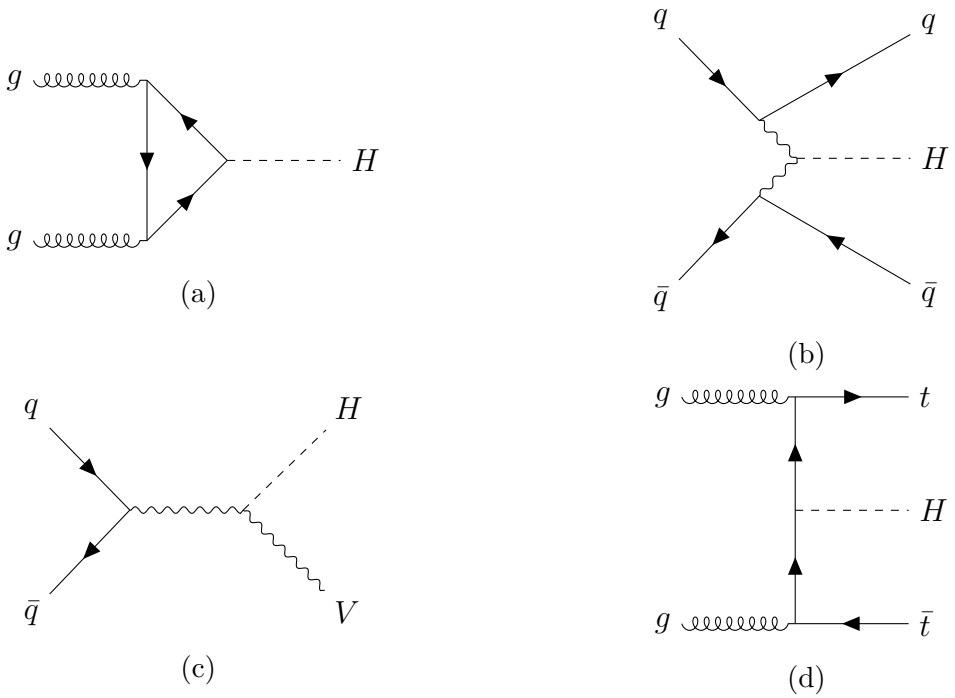


Figure 2.4: Feynman diagrams for the main Higgs production modes at the LHC: (a) gluon-gluon fusion (ggF), (b) vector boson fusion (VBF), (c) associated production with a vector boson (VH) and (d) associated production with a top quark pair ($t\bar{t}H$).

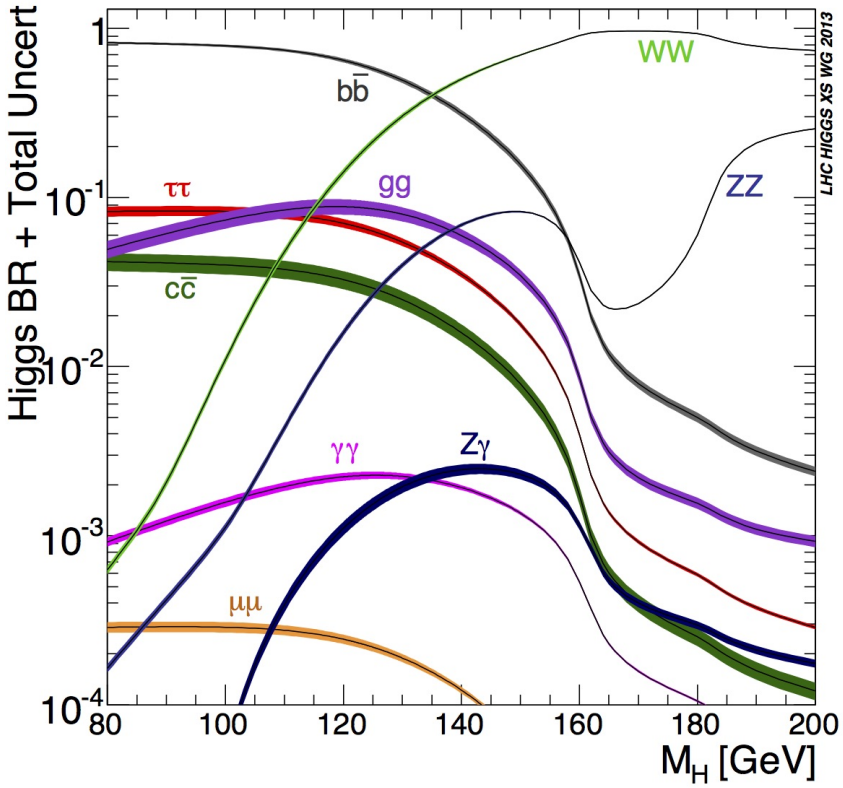


Figure 2.5: The SM Higgs boson decay branching ratios and widths as a function of Higgs mass [13].

2.3 Fermions

Fermions are spin 1/2 particles that make up matter we see in our universe and are described by $\mathcal{L}_{Fermions}$ part of the Lagrangian. They follow the Pauli exclusion principle which states that no two spin 1/2 particles may occupy the same state (e.g. electron orbitals in an atom) in a quantum system. Typical matter (comprising of atoms) is made up solely out of the first generation of fermions (up and down quarks, electrons forming protons and neutrons). The three generations of fermions have the same properties in terms of colour, weak and electromagnetic charges and only vary in terms of their masses. Each fermion has a specific flavour associated with it, with only charged weak interactions (mediated by the $W^{\pm}$ boson) allowing a fermion to change from one flavour to another. The Cabibbo-Kobayashi-Maskawa (CKM) matrix [14, 15] contains the transition probabilities for quarks from one flavour to another. There are three mixing angles that represent the rotation of the mass eigenstate vector space into the weak eigenstate vector space. The mixing angels can be defined in terms of the ratio of the transition probabilities. The CKM matrix also accounts for the observed CP violation in weak decays, introducing a fourth free parameter.

2.3.1 Top quark

The top quark is the heaviest particle in the SM and is part of the third generation of fermions. It was discovered in 1995 at the Tevatron by the CDF [16] and D0 [17] collaborations. Because it is so heavy (m_t = 173.1 GeV) it is extremely unstable and decays almost instantaneously ($\tau \sim 10^{-25}$ s) and doesn't undergo hadronisation ($\tau \sim 10^{-22}$ s), which allows the study of a bare quark. Its typical production mode is a pair of top quarks ($t\bar{t}$) via gluon-gluon fusion and decays into a b quark and W boson. Given the larger mass of the top quark, the Higgs boson will (almost) never decay to the $t\bar{t}$ final state. Measuring the branching ratios of the Higgs is one way to access the Yukawa couplings. However, since this is not possible for the top Yukawa coupling the best way to access this parameter is through the study of the $t\bar{t}H$ cross-section.

2.4 Strong Interaction

The strong nuclear interaction is described by QCD and governs the interaction between the quarks and gluons. The colour charge C is the conserved quantum number with there being a total of three colour charges. The theory follows the $SU(3)_C$ gauge symmetry with there being a total of 8 generators of the group. Quarks carry one colour charge each with the interaction being mediated by the massless gluon boson. The strong coupling "constant" α_s depends on energy (so is not a constant) and describes the strength of the

interaction. The gluons carry two colour charges and as a result can interact with themselves. It is due to this self interaction that quark confinement exists where particles with colour charge bind together to form colour neutral states. At low energy the gluons self interaction creates an anti-screening effect which amplifies the strength of the interaction.

2.5 Parameters of the Standard Model

The free parameters of the SM (19 + 7 in total) can be defined in some different ways. Either the 9 fermion masses or their Yukawa couplings $y_f = \sqrt{2}m_f/\nu$ can be used. The three mixing angles (θ_{12}, θ_{23}, θ_{13}) from the CKM matrix as well δ the CP violating phase. The three gauge couplings (g_1, g_2, g_3) can be used with alternatives for g_1, g_2 being the weak mixing angle $\tan\theta_w = \frac{g_1}{g_2} = \frac{m_w}{m_z}$ and the fine-structure constant $\alpha = \frac{1}{4\pi}\frac{(g_1 g_2)^2}{g_1^2+g_2^2}$ of QED. The final 3 parameters are the QCD vacuum angle θ_{QCD}, the Higgs vacuum expectation value ν and the Higgs mass m_H.

The 7 extra parameters related to neutrinos are the 3 neutrino masses, the 3 flavour mixing angles and δ_{CP} which is the CP violating term. This information is contained in the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [18, 19] which is analogues to the CKM matrix and describes neutrino mixing.

2.6 Beyond the Standard Model

There are a number of phenomena not yet explained by the SM with extensions to the theory being referred to as Beyond the Standard Model (BSM) theories. The most obvious of which is gravity for which no quantum theory currently exists and is best described by general relativity. Other than that there is no a priori reason in the SM for the masses of the fermion generations to have such different scales. It also doesn't explain dark matter nor dark energy as well as the matter-antimatter asymmetry.

2.7 $t\bar{t}H$ and the top Yukawa coupling y_t

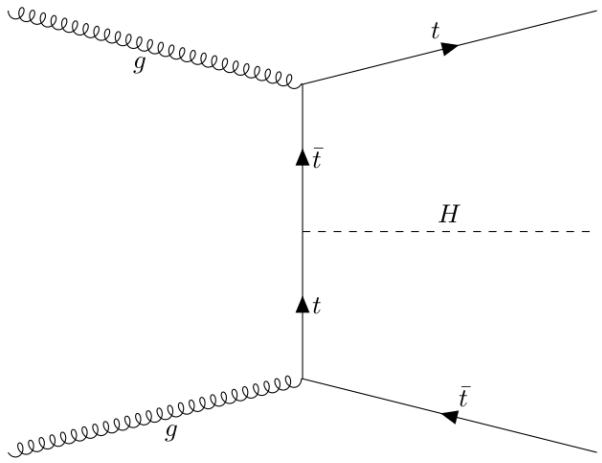
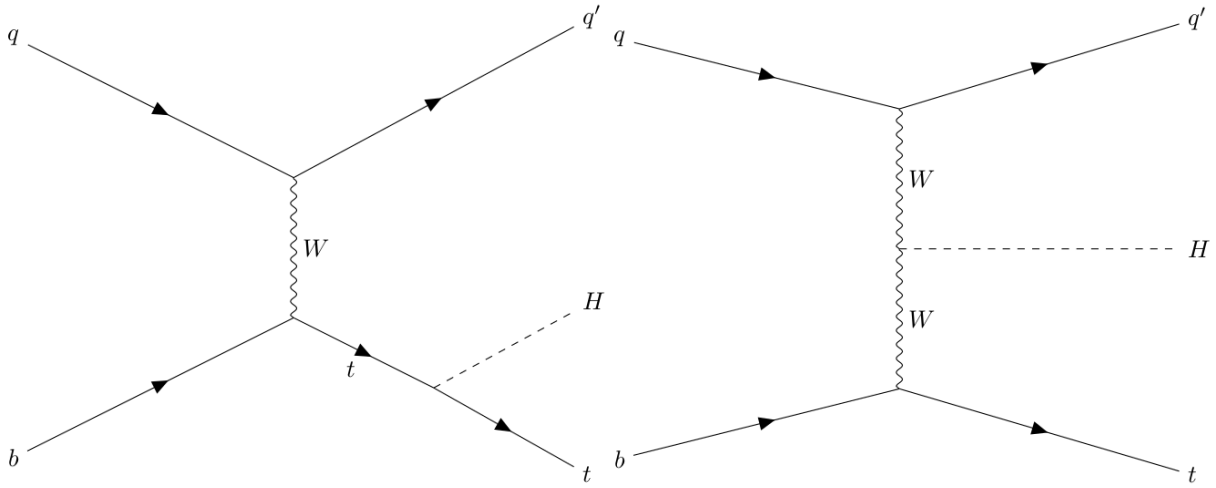


Figure 2.6: Example of a tree-level Feynman diagram of the $t\bar{t}H$ t-channel process.

The Higgs top Yukawa coupling y_t is one of the fundamental parameters of the standard model and needs to be measured. It is directly accessible at tree level with $t\bar{t}H$ (Figure 2.6) and tH (Figure 2.7) or indirectly with a virtual top quark loop e.g $H \to \gamma\gamma$ (Figure 2.8) corresponding to the Higgs-top vertex in the Feynman diagram amplitude. The $t\bar{t}H$ process was discovered by both ATLAS [20] and CMS [21] in 2018 by combining various Higgs decay channels.



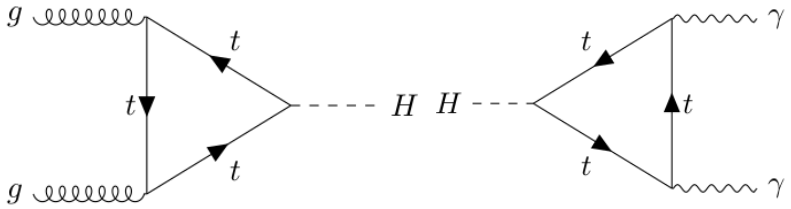
(a) tH t-channel process with the Higgs coupling to the top quark.

(b) tH t-channel process with the Higgs coupling to the W boson.

Figure 2.7: Leading order Feynman diagrams for t-channel tH production in the 5 Flavour Scheme.

Measuring the $t\bar{t}H$ cross-section is a direct way to access and measure the Higgs top Yukawa coupling y_t. The $t\bar{t}H$ process is preferable for measuring the magnitude $|y_t|$

because it has a higher predicted cross-section than tH. On the other hand tH gives access to the sign of the coupling due to the interference of the W and top couplings.



(a) GGF Higgs production mediated through a virtual top quark loop.

(b) $H \to \gamma\gamma$ decay mediated through a virtual top quark loop.

Figure 2.8: Examples of indirect y_t couplings through virtual top quark loops.

The predicted cross-sections of $t\bar{t}H$ and tH can be seen in Figure 2.9, where the SM Lagrangian has been extended such that the effective Lagrangian for Higgs-top quark interactions below the EWSB scale is

$$\mathcal{L}_0^t = -\frac{y_t}{\sqrt{2}}\bar{\psi}_t(\cos(\alpha)\kappa_{Htt} + i\sin(\alpha)\kappa_{Att}\gamma_5)\psi_t X_0.$$

Where X_0 labels a generic spin-0 particle (i.e the Higgs) with CP-violation couplings, α is the CP-mixing phase, κ_{Htt}, κ_{Att} are real dimensionless rescaling parameters. With the above approach a CP-even case would be $(\cos(\alpha) = 1, \sin(\alpha) = 0)$, CP-odd case would be $(\cos(\alpha) = 0, \sin(\alpha) = 1)$ and the SM case $(\cos(\alpha) = 1, \kappa_{Htt} = 1)$. For more details of the approach taken see [22].

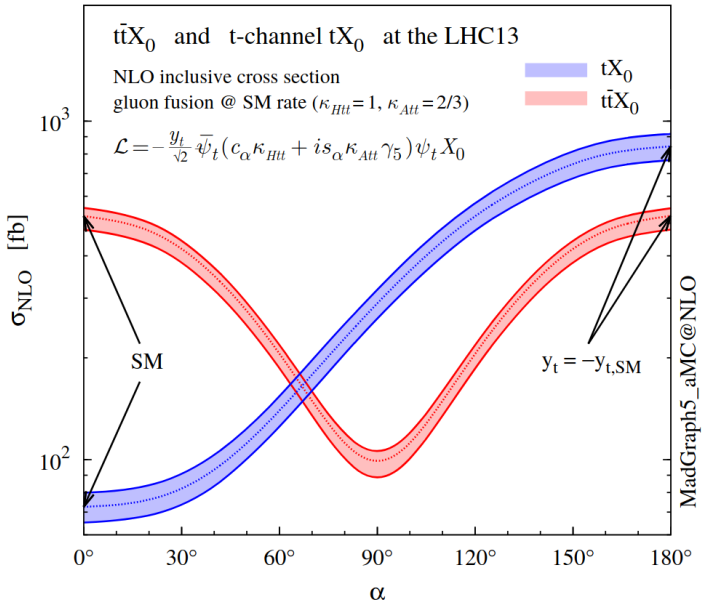


Figure 2.9: The NLO cross-sections (with scale uncertainties) for $t\bar{t}X^0$ and t-channel tX_0 productions at the 13-TeV LHC as a function of the CP-mixing angle α, where κ_{Htt} and κ_{Att} are set to reproduce the SM GF cross section for every value of α [22].

From Figure 2.9 we can see that $t\bar{t}H$ is not sensitive to the sign of the coupling (i.e $y_t \to -y_t$) because it depends quadratically on y_t. However, it is sensitive to a CP-odd / CP-mixed scenario in which case its cross-section would be reduced. We can also see that tH is clearly sensitive to the sign of y_t due to the interference between the diagrams where the Higgs couples to the top quark and W boson as can be seen in Figure 2.7. The above also highlights the interplay between the two channels ($t\bar{t}H$ and tH) at different values of α and how important it is to separate the two as $t\bar{t}H$ alone is also not sensitive to $\alpha \to \pi - \alpha$ while tH is predicted to have an order of magnitude increase in cross-section in that case.

Decay channel	Approximate branching ratio
$H \to b\bar{b}$	5.8×10^{-1}
$H \to W^+W^-$	2.1×10^{-1}
$H \to \tau^+\tau^-$	6.3×10^{-2}
$H \to ZZ$	2.6×10^{-2}
$H \to \gamma\gamma$	2.3×10^{-3}
$H \to Z\gamma$	1.5×10^{-3}
$H \to \mu\mu$	2.2×10^{-4}

Table 2.1: Table of H decays and corresponding branching ratios, values taken from [13].

The $H \to b\bar{b}$ is the dominant decay mode (see Table 2.1). However, it is also one of the hardest decay channels to analyse because it does not give a clear signal to trigger on like $H \to \gamma\gamma$ or $H \to ZZ \to 4l$. The $H \to \gamma\gamma$ decay channel, whilst being one of the smallest branching ratios, has actually provided one of the cleanest signals so far and drove the initial Higgs discovery.

There have been a few dedicated $t\bar{t}H \to b\bar{b}$ results by both ATLAS and CMS. In the initial observations these dedicated results were combined with other decays to reach the required 5σ observed significance. These used the collaborations respective 2016 Run-2 data subsets with ATLAS using $36.1fb^{-1}$ [23] and CMS using $35.9fb^{-1}$ [24]. Since then both collaborations have released updated results using larger data sets. ATLAS released a full Run-2 result using $139fb^{-1}$ [1] and is the analysis in Section 8.2 where my contributions to the boosted region are detailed. CMS released a 2016/17 combined result using $77.4fb^{-1}$ [25].

Chapter 3

The ATLAS detector and the LHC

The LHC based at CERN is the worlds highest energy particle accelerator built to date. It re-uses the same 27 km tunnel used by LEP and houses four LHC experiments: ATLAS, CMS, ALICE and LHCb. They are spaced along the ring at the four primary interaction points. Protons bunches are accelerated and collided together at the primary interaction points at $\sim$ 14 TeV every 25ns. This is achieved through several sub-accelerators which sequentially accelerate the bunches till they are fast enough to be injected into the primary LHC ring. The bunches are bent into curved trajectories through the use of superconducting magnets which can produce magnetic fields of up to 7.7 T. The primary limiting factor for the LHC, in the achievable centre of mass energy is the magnet technology and its achievable field strength. This limits the inward acceleration of the particles required for a given radius. LEP initially used leptons (electrons/positrons) which are lighter and so limited by synchrotron radiation. Leptons are more precise but achieve a lower centre of mass energy than hadrons. Thus switching to hadrons (protons) provides a natural progression whilst making use of the same tunnel which due to their greater mass will naturally produce a higher centre of mass energy. The inherent downside to this is that the exact momentum fraction in a given hard scattering event is unknown due to the partons of the protons being the actual collision constituents. This leads to hadron colliders having inherently larger uncertainties than lepton colliders but allows the probing of significantly higher energies. As such lepton colliders are beneficial for precision measurements whilst hadron colliders are advantageous for discovering new particles. This was the primary goal of the LHC with the discovery of the Higgs boson in 2012.

The LHC experiments are organised around the detector and the beam. Collaborations of physicists work on designing the detectors, operating them during data taking and finally analysing the data. Since experiments with general purpose detectors such as ATLAS and CMS are supposed to verify each others results, physicists working on data analysis are only allowed to be members of one LHC experiment. The LHC and its experiments are organised into data taking run periods, with breaks in-between to allow

for maintenance of both the ring and the various detectors. This allows for the continuous ongoing development of both the accelerator ring and detectors to be installed. This chapter is organised into two subsections. The first section focuses on the LHC run periods, specifically Run-2. The second section explains the ATLAS detector in its Run-2 configuration.

3.1 LHC runs

The LHC currently has four planned operational runs, with the first two (Run-1 and Run-2) having successfully been completed as of 03/12/2018. Run-3 commenced on 05/07/2022 and was delayed due to the COVID pandemic. Run-1 took place between 2009 and 2013 and operated at centre of mass energies of $\sqrt{s} = 900$ GeV, $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. Run-2 started approximately two years later in 2015 and gathered data till the end of 2018 and achieved a centre of mass energy of $\sqrt{s} = 13$ TeV and a bunch spacing every 25ns. The data taking periods in ATLAS refer to 2015/16, 2017 and 2018 with the matching Monte Carlo campaign labels being mc16a, mc16d and mc16e respectively. The Monte Carlo campaigns are designed to match the run conditions specific to those periods including luminosity and detector radiation damage. Luminosity L is a measure of the number of events being produced at a given moment and is defined as:

$$L = \frac{kN^2}{4\pi s a_h a_v},$$

where s is the bunch spacing, k is the number of bunches per beam, N is the number of protons per bunch and a_h and a_v are the effective beam sizes in the horizontal and vertical directions. Integrating this quantity over the whole data taking period defines the integrated luminosity and is measured in inverse femtobarns (fb^{-1}) with $1\text{fb} = 10^{-39}\text{cm}^2$.

The analyses performed here utilise the full Run-2 ATLAS data set of 139fb^{-1} at $\sqrt{s} = 13$ TeV. This is because while a total of 156fb^{-1} was delivered by the LHC, 147fb^{-1} was recorded by ATLAS of which 139fb^{-1} was usable for physics, where all the subsystems were fully operational. The breakdown of delivered and recorded luminosity by year for ATLAS can be seen in Figure 3.1 with the total luminosity for Run-2 shown in Figure 3.2 where the delivered, recorded and good for physics data are detailed.

For a given bunch crossing in ATLAS the chance for more than one pp collision per crossing is rather high and leads to an effect called pile-up (μ). This increases with luminosity and has the effect of increasing "noise" making it harder to reconstruct the interesting hard scattering events. With the original design luminosity of the LHC ($10^{34}\text{cm}^{-2}s^{-1}$) being exceeded in 2016 and a maximum luminosity of $2.1 \times 10^{34}\text{cm}^{-2}s^{-1}$ achieved in 2018 led to an average pileup, $\mu = 33.7$ compared to the expected value of $\mu = 23$. This difference

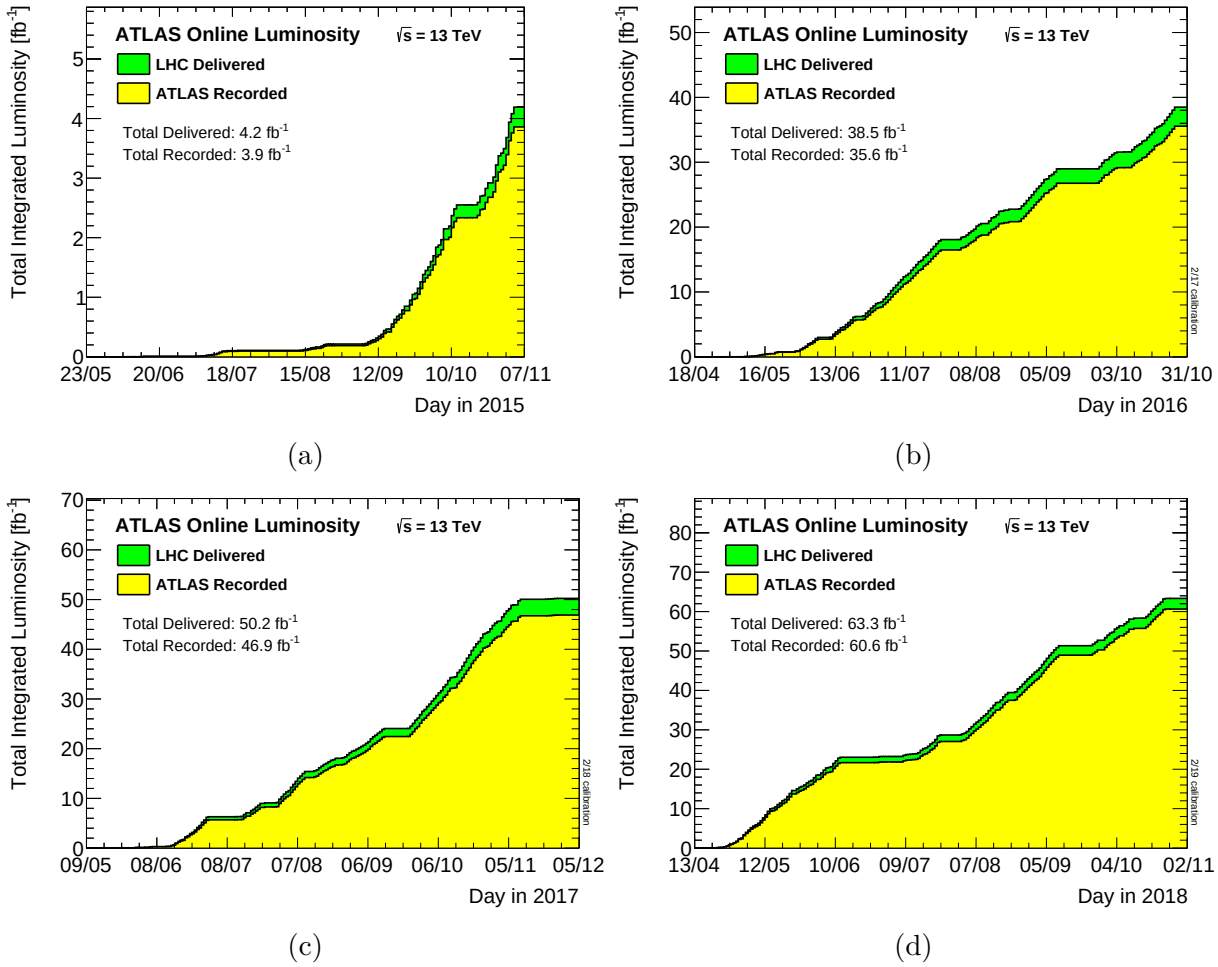


Figure 3.1: The luminosity delivered by the LHC and recorded by ATLAS for each year during Run-2 [26].

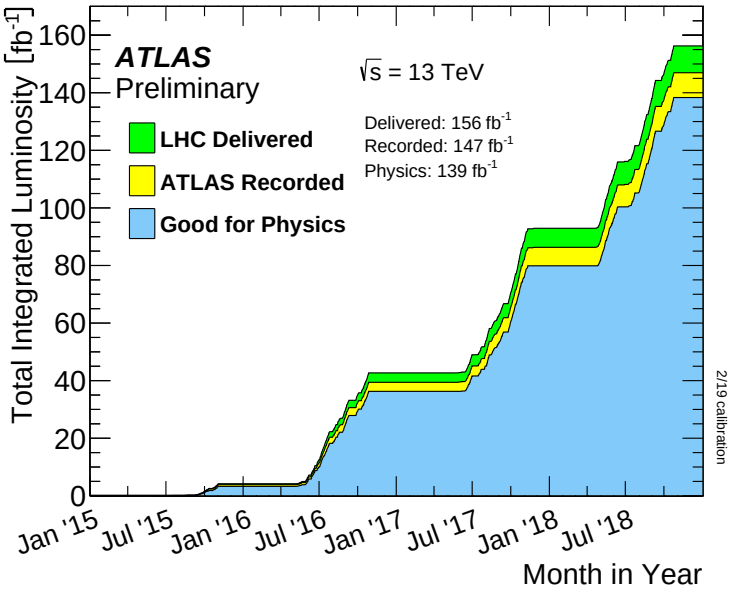


Figure 3.2: The total luminosity delivered by the LHC, the total luminosity recorded by ATLAS and the total luminosity of data usable for physics analysis for Run-2 [26].

has primarily been dealt with through continuous improvements in reconstruction algorithms. A breakdown of the average number of interactions per crossing and associated μ for each run period and the whole of Run-2 can be seen in Figure 3.3.

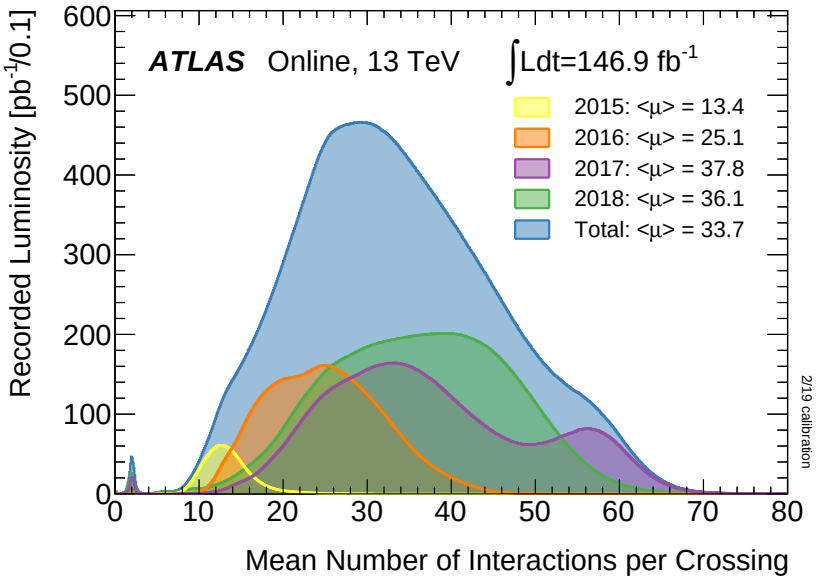


Figure 3.3: The average number of pp interactions per bunch crossing in ATLAS for each given year and the whole of Run-2 [26].

Run-3 is now ongoing with a centre of mass energy of $\sqrt{s} = 13.6$ TeV which maximises what is possible with the current magnet technology. The plan is to deliver $\sim 280\text{fb}^{-1}$ [27] which will approximately triple the ATLAS data set available for physics analysis. The next run (Run-4) plans to be a high luminosity (HL) run where extra performance is squeezed out of the available tunnel not by increasing the centre of mass energy but by increasing the bunch sizes. This will lead to an increase in luminosity and by extension an increase in pile-up. To mitigate this a substantial redesign of the ATLAS detector is ongoing with an all new fully silicon tracker being implemented called ITk. As part of service work for the ATLAS collaboration some modelling for the wire-bonding process during manufacturing for the ITk strips was performed. This is detailed in Chapter 4.

3.2 The ATLAS detector

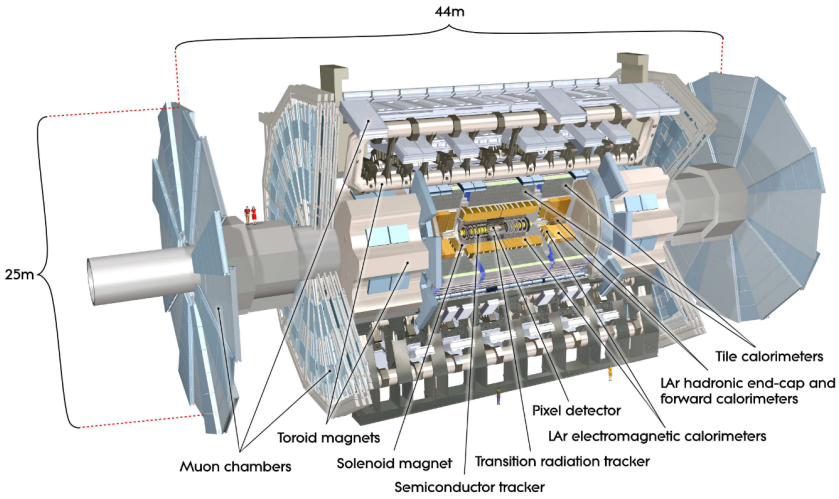


Figure 3.4: A cutaway of the ATLAS detector [28].

The ATLAS detector [28], depicted in Figure 3.4, is one of the two general purpose detectors at the LHC, with the other being CMS. The name ATLAS is derived from A Toroidal LHC ApparatuS and the experiment can probe both proton-proton collisions as well as heavy ion collisions. Its primary purpose is to identify and reconstruct particles that emerge from proton-proton collisions and its initial design was heavily influenced by the then theorised and now discovered Higgs boson. To do this the particles' trajectories and energies are recorded from which basic kinematic variables can be derived. A particles' charge and momentum can be inferred from the bending direction and curvature of its trajectory. This information is then used for both particle identification and reconstruction. The coordinates used are defined with the nominal interaction point (IP) as the origin. The z-axis runs parallel to the beam line with the x-y plane being transverse to it. The azimuthal angle ϕ is measured around the beam line and the polar angle θ is measured from the beam line. Due to the particles being relativistic in the direction parallel to the beam pseudorapidity $\eta = -\ln\tan(\theta/2)$ is used which is an approximation of the Lorentz invariant quantity rapidity for $m \gg p_T$.

The detector is comprised of several subsystems akin to an onion with each specialised at measuring different properties. The main subsystems listed in increasing distance from the IP are: Inner Detector (ID), Calorimeters, Muon Spectrometer (MS) and the magnet systems. On-top of these the trigger and data acquisition systems (TDAQ) are integral for maximising the amount of useful data collected and are further split into three levels: L1, L2 and event-filter. Due to both the Run-2 detector design and data taking having been completed prior to any of the studies described here commencing, only a brief outline will be given of each subsystem and TDAQ as there was no scope to be involved in these steps.

However, details of the ITK strips for ATLAS ID upgrade designed for Run-4 HL-LHC will be given in Chapter 4 as part of service work for the ATLAS collaboration to qualify for ATLAS authorship. The detector has a cylindrical geometry with components divided into barrel and endcap, where barrel components typically have strip geometries and are positioned parallel to the beam while endcap components typically have petal geometries and are positioned perpendicular to the beam.

Different particles interact differently with each subsystem and so can be identified through combining their unique signatures from each subsystem. A depiction of this can be seen in Figure 3.5.

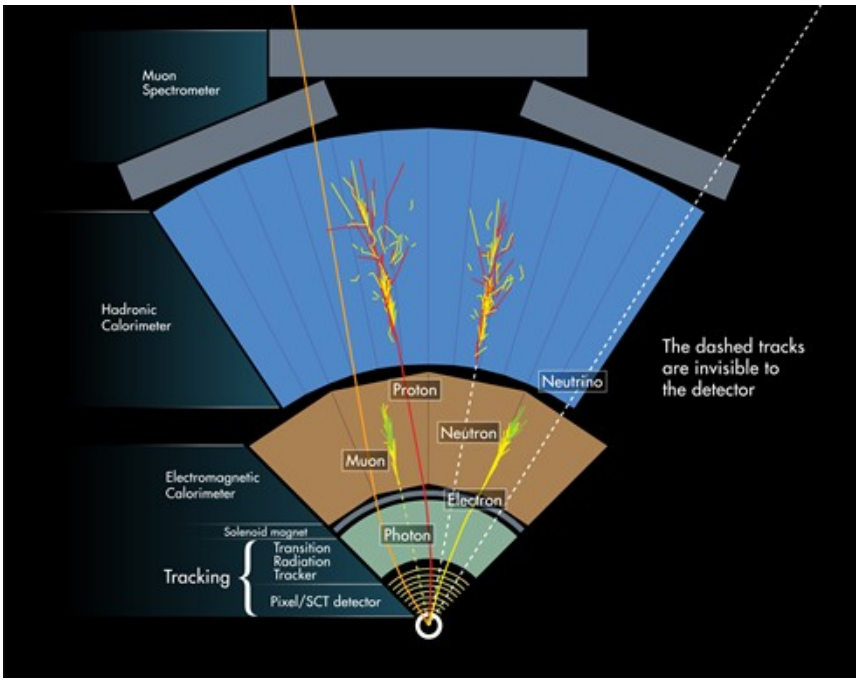


Figure 3.5: An illustration of how different particles interact with the various ATLAS sub-detectors [29].

3.2.1 Inner Detector (ID)

The Inner Detector (ID) [28], depicted in Figure 3.6, is the subsystem closest to the beam and is used to measure the trajectories of charged particles. As it is closest to the interaction point it operates in the busiest particle environment and so requires the highest resolution out of all the subsystems. It is comprised of three separate technologies: Pixel, Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT). The whole ID is immersed in a 2T magnetic field from the magnet systems. The tracks are used to measure the momentum of the charged particles (due to the bending from the magnetic field) as well as primary and secondary vertices.

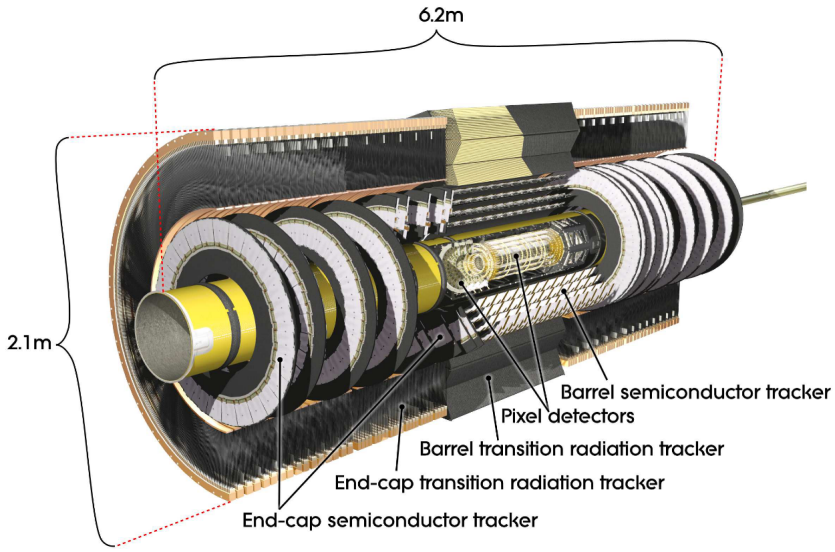


Figure 3.6: A close-up cross section of the ATLAS ID tracking subsystems [28].

Pixel detector

The Pixel detector is the closest to both the beam pipe and IP. It consists of 4 barrel (including the IBL) and 3 endcap layers of silicon sensors providing up to $|\eta| < 2.5$ coverage with tracks hitting at least 3 layers, see Figure 3.7. The insertable b-layer IBL [30] was added between Run-1 and Run-2 to improve b-hadron identification and was inserted at a radius of $\approx$ 33mm [30]. The IBL has pixels of size $50 \times 250\mu\text{m}^2$ with the rest of the pixels, in the other layers, having a size of $50 \times 400\mu\text{m}^2$. The spatial resolution achieved in $R - \phi$ (beam perpendicular) is 10μm and in z is 115μm (beam parallel), with the IBL having spatial resolutions of 8μm and 75μm respectively.

Semiconductor Tracker (SCT)

The next set of sensors are the Semiconductor Tracker (SCT) which also uses silicon sensors but with a notable difference in readout. The sensors consist of strips with each hit giving a 1 dimensional measurement. Because of this each SCT layer consists of sets of stereo strips to measure both coordinates, with a space-point requiring 2 hits on either side of the layer. The SCT consists of 4 barrel layers and 9 endcap disks giving it a $|\eta| < 2.5$ coverage. This ensures that there are at least 4 space-points (8 hits) per track in this region, see Figure 3.7, with a spacial resolution of 17μm and 580μm in the transverse and longitudinal directions respectively.

Transition Radiation Tracker (TRT)

The Transition Radiation Tracker (TRT) is the outermost part of the ID. It uses 4mm diameter straw tubes filled with a gas mixture (predominantly Xenon) and only provides

measurements in $R - \phi$. While it only has $|\eta| < 2.0$ coverage, see Figure 3.7, it does provide a large number of hits per track ≈ 36 with a spatial resolution of 130μm. Its primary use is to aide in electron identification by aiding in tracking and identifying the resultant bremsstrahlung radiation.

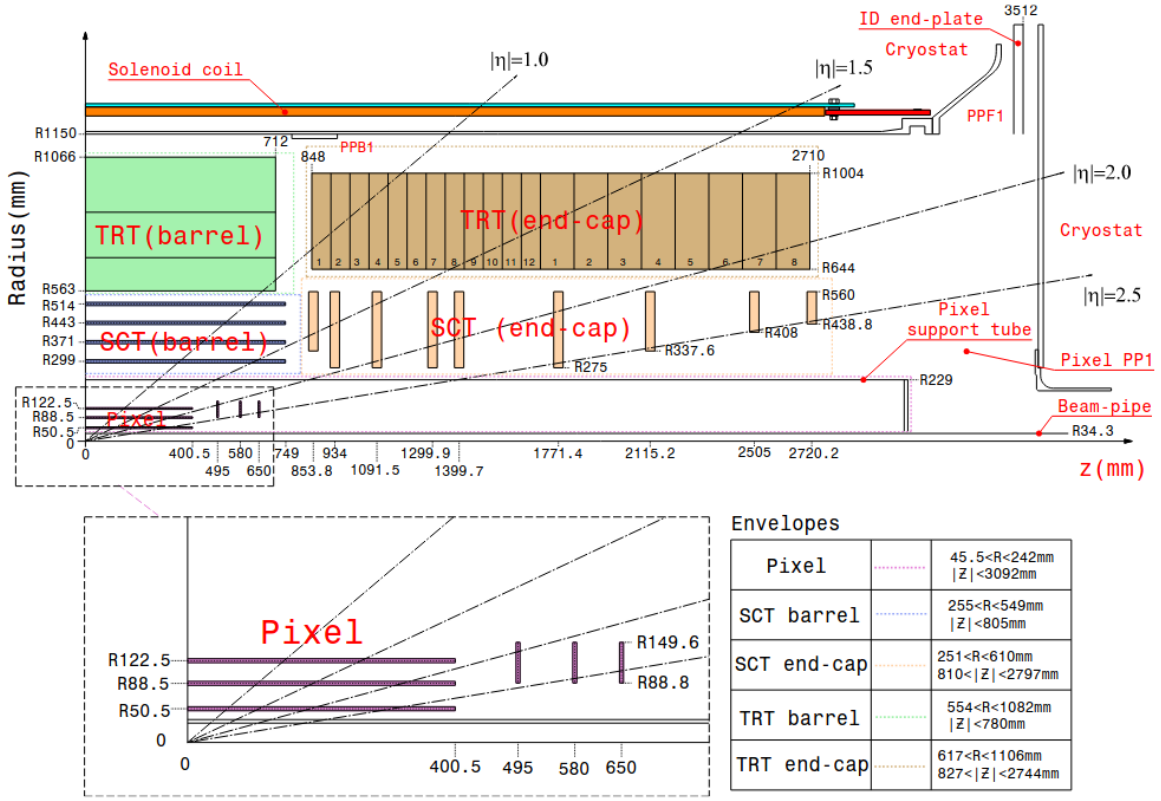


Figure 3.7: A schematic layout of the ID tracking sub-detectors for both barrel and end-cap sections depicting various η ranges [28].

3.2.2 Calorimeters

The next layer of the ATLAS onion after the inner tracking detectors are the calorimeters [28] that measure the energies of particles. They do this using a dense metal plate called an absorber that effectively has a high interaction cross-section with incident particles. Once the incident particles hit the absorber they causes a resultant shower of particles that slowly propagates through the sensing elements dissipating energy thus allowing for the measurement of the incident particles energy. There are two distinct shower types electromagnetic (EM) showers, that are induced by high energy photons/electrons/positrons and cause a cascade via pair production and bremsstrahlung, and hadronic showers induced by high energy hadrons and have more processes inducing the resultant shower.

An important consideration is to design the calorimeters to be thick enough to capture

the whole cascade of particle showers and limit punch-through of charged particles into the muon subsystems. Because of the different nature of the EM and hadronic showers there are two distinct calorimeter systems designed for each respective particle shower type. The combined calorimeter systems are designed such that there is almost no punch-through of ionising radiation with the exception of the minimum ionising muons which are detected by the dedicated muon systems. A depiction of the ATLAS calorimeter systems can be seen in Figure 3.8.

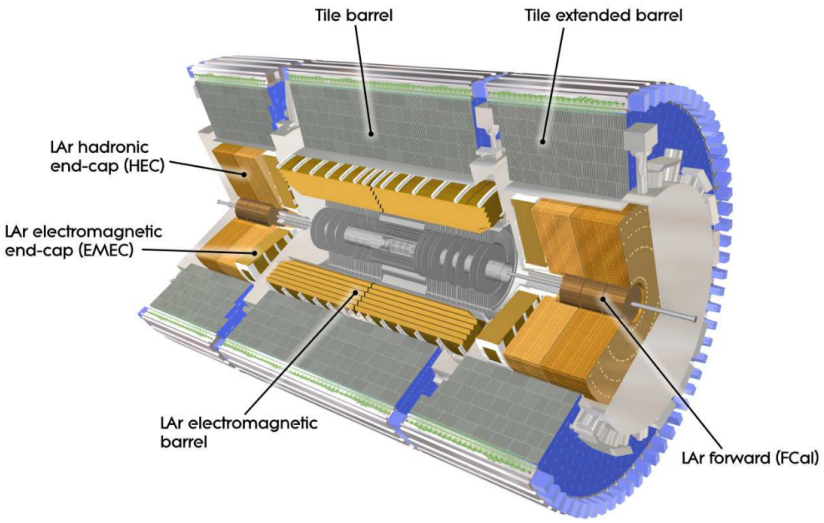


Figure 3.8: A cutaway of the ATLAS detector depicting the various calorimeter systems and sections [28].

LAr Electromagnetic Calorimeter

The first calorimeter layer encountered by incident particles is the EM calorimeter which uses lead as the material for the absorber plates and liquid argon as the sensing material. It is configured to provide the highest resolution ($\approx 0.025 \times 0.025$ in $\Delta\eta \times \Delta\phi$) for $|\eta| < 2.5$, inline with the ID, but covers the region up to $|\eta| < 3.2$. It consists of both barrel and endcap sections as can be seen in Figure 3.8. Due to the solenoid magnet's position in front of the EM calorimeter they share a common vacuum vessel thereby eliminating two extra vacuum walls. For the high precision region ($|\eta| < 2.5$) the calorimeter has sensor segments three layers deep. The rest of the calorimeter only has sensor segments two layers deep and a coarser granularity ($\approx 0.1 \times 0.1$ in $\Delta\eta \times \Delta\phi$). In addition there is a pre-sampler that covers the $|\eta| < 1.8$ region that consists of an active LAr layer and is used to correct the energy lost by electrons and photons prior to the calorimeter.

Hadronic Calorimeters

Next in line are the hadronic calorimeters which consist of tile calorimeters in the barrel, LAr hadronic end-cap calorimeter and the LAr forward calorimeter. The tile calorimeter

use steel as the absorber and scintillating tiles as the active medium. It consists of two sections: barrel covering $|\eta| < 1.0$ and extended-barrel covering $0.8 < |\eta| < 1.7$. Similarly it has three layers of sensing tile segments with the initial two layers having a granularity of 0.1×0.1 in $\Delta\eta \times \Delta\phi$ and 0.2×0.1 in the final layer. The LAr hadronic endcap calorimeter, similarly to the EM calorimeter, uses lead as the absorber and LAr as the sensing element and shares the cryostat of the EM calorimeter. It covers a region of $1.5 < \eta < 3.2$ in order to overlap both with the tile calorimeters in the barrel and the LAr forward calorimeter. It has a granularity of 0.1×0.1 for $1.5 < |\eta| < 2.5$ and 0.2×0.2 for $2.5 < |\eta| < 3.2$.

The LAr forward calorimeter

The LAr forward calorimeter extends the calorimeter coverage up to $|\eta| < 4.9$ and is integrated into the endcap cryostat. Due to the high radiation levels present due to neutron albedo the forward calorimeter is recessed by 1.2 m with respect to the EM calorimeter, see Figure 3.8. As a result and to prevent punch through a high density design is necessitated. It consists of three modules with the first made of copper and optimised for electromagnetic showers. Tungsten is used for the other layers and optimised for hadronic showers. Each module forms a metal matrix of electrodes with small gaps of LAr acting as the sensing medium. This design, while more complex, avoids problems caused by ion build up.

3.2.3 Muon Spectrometer (MS)

The outermost layer of the ATLAS onion is the muon spectrometer (MS) [28] and is designed to measure the momentum of charged particles that pass through the previous ATLAS-subsystems (ID and calorimeter), which predominantly tend to be muons. Similarly to the ID the momentum and charge are inferred from the bending of the muons trajectories present due to the magnetic field. The trajectories are reconstructed as tracks using hits in the various MS-subsystems. The four main MS-subsystems are: Monitored Drift Tubes (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). With the MDT and CSC providing precision tracking and the RPC and TGC providing fast tracking used by the trigger system. An illustration of the MS-subsystems can be seen in Figure 3.9.

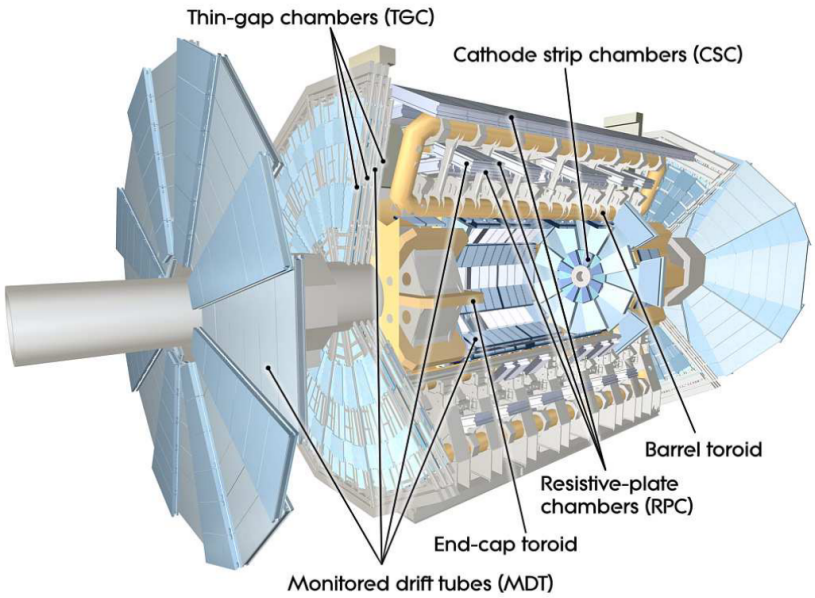


Figure 3.9: A cutaway of the ATLAS detector depicting the various MS-subsystems [28].

Monitored Drift Tubes (MDT)

Precision tracking is provided primarily by the MDT chambers arranged in three layers (for both barrel and endcap). It covers the $|\eta| < 2.7$ region with the exception of the inner most endcap layer which only covers $|\eta| < 2.0$. Each MDT chamber is made up of 3-8 layers of drift tubes giving a chamber resolution of 35 μm. A schematic of the MDT chambers can be seen in Figure 3.10 and are green for barrel and teal for endcap sections.

Cathode Strip Chambers (CSC)

In the innermost endcap layer, due to the higher rate and time resolution requirements, the CSC provides precision tracking for $2.0 < |\eta| < 2.7$ region. These are multi-wire proportional chambers with cathodes segmented as strips in orthogonal directions with a chamber resolution of 40 μm. The CSC can be seen in yellow in Figure 3.10.

Trigger Chambers

In order to be able to trigger on muons, information on their tracks must quickly be sent to the L1 trigger logic. In addition the muon trigger system aims to fulfil four specific criteria: provide bunch-crossing identification, provide well defined p_T thresholds, provide robustness from random hits due to n/γ-background and provide a second coordinate measurement in the non-bending ϕ-projection to complement the precision tracking chambers. For the barrel this is provided by Resistive Plate Chambers (RPC) in the $|\eta| < 1.05$ region and provide a time resolution of 1.5ns, in white in Figure 3.10. In the endcap Thin Gap Chambers (TGC) are used covering the $1.05 < |\eta| < 2.4$ region and have a time resolution

of 4ns, in purple in Figure 3.10. Both chamber types have a response time of 15-25ns which enables them to be used to tag the beam crossing.

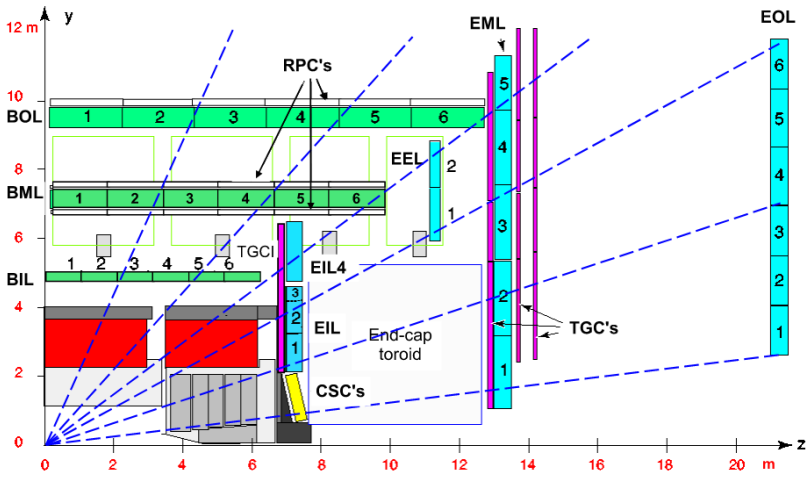


Figure 3.10: A schematic view of the MS and its subsystems [28].

3.2.4 Magnet Systems

The ATLAS detector employs a set of four superconducting magnets [28] consisting of one solenoid magnet surrounding the ID, and three toroid magnets (with one dedicated to the barrel and one for each endcap) that provide the magnetic field for the MS. These can be seen in Figure 3.11.

The solenoid surrounding the ID provides a 2T magnetic field for bending charged particles. Due to its position in the detector careful consideration was given so as not to introduce too much material in-front of the calorimeters. To achieve superconductivity it is cooled to 4.5K with liquid helium.

The barrel and endcap toroids provide the magnetic field for the MS, providing magnetic fields of 0.5T and 1T respectively for their given regions. They both consist of eight coils, with the barrels coils each being housed in their own vacuum vessel. Similarly they are all cooled with liquid helium to around 4.5K to achieve superconductivity.

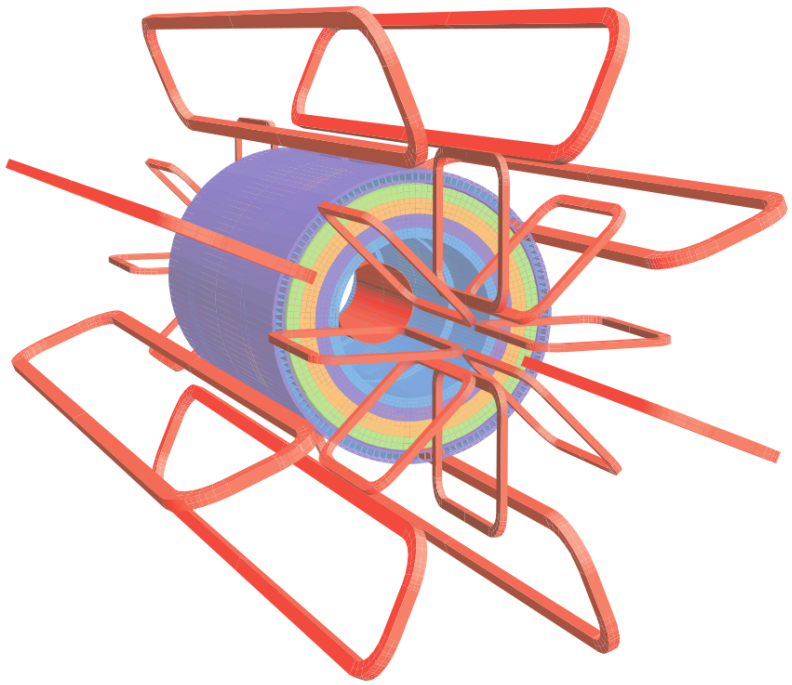


Figure 3.11: Geometries of the magnet systems [28].

3.2.5 Trigger and data acquisition (TDAQ)

Since all the LHC experiments share the same accelerator they all receive the same amount of bunch crossings, dictated by the bunch spacing, which is 25ns. That is an equivalent rate of $\sim$ 40 MHz. However, not every bunch crossing will yield an interesting hard scattering event. Moreover, it is simply not possible to initially store, analyse and then discard such a large amount of data. Thus a way of quickly determining events of interest in real-time is required. This is achieved through the use of a "trigger" which selects interesting events to store through some pre-defined criteria. This is done in two stages with the Level-1 (L1) trigger reducing the event rate to $\sim$ 100 KHz and the High-Level Trigger (HLT), which is the combination of the Level-2 (L2) trigger and event filter, reducing the event rates to $\sim$ 3.5 KHz and $\sim$ 200 Hz respectively [28].

L1 trigger

The L1 trigger is based on custom electronics and uses information from the muon trigger systems (RPC and TGC) as well as the calorimeter. It searches for high p_T muons, electrons, photons, jets, hadronically decaying taus and p_T^{miss}. The overall decision is made by the Central Trigger Processor (CTP) which selects events if they pass one of the 256 programmable options of the trigger "menu". Each option is essentially just a combination of criteria on the various objects it is trying to select for. Once an event is accepted by the L1 trigger it is passed to the slower HLT where more complex event

selection can be performed using conventional CPUs. In the meantime event information is temporarily stored while the HLT makes a decision. In addition the L1 trigger identifies Regions of Interest (RoI's) in terms of η-ϕ space that contain combinations of features that meet their given criteria which are then passed on to the HLT.

HLT trigger

The HLT uses more conventional CPUs to reconstruct events using all the information available to the L1 trigger as well as tracking information from the ID. Overall it consists of many algorithms organised sequentially in increasing complexity and CPU intensity. This has the obvious benefit that should a candidate fail to pass at an early stage, unnecessary CPU resources are not wasted on reconstructing it in greater detail. The HLT consists of two parts, the L2 trigger and the event filter. The L2 trigger is initially seeded with information from the RoI's, determined by the L1 trigger, using the coordinate, energy and signature-type information to reduce the amount of data. The event filter then makes the final selection on which events are permanently stored. To that end events passing the L2 are initially fully reconstructed with the event builder, from the data acquisition systems cache, and then filtered with the use of offline analysis techniques.

Chapter 4

ITk Strip Module Metrology

My Qualification task was focused on helping with ITk Strip module metrology for the ITk upgrade. The current ID is designed to operate for 10 years and would have a number of limitations operating in the HL-LHC. There would be bandwidth saturation for the Pixels and SCT, there would be increased occupancies in the TRT and SCT and there would be significant radiation damage to the Pixels and SCT.

The ITk will be an all silicon design with a complex geometry to minimise surface area and cost. It will cover $|\eta| < 4.0$ with 5 Pixel barrels and 4 Strip barrels, see Figure 4.1. The strip end-caps will be made from 6 disks each side. While the Pixels, at higher η, will initially be laid out in a complex geometry at various inclines and then in rings. Together the Pixels and Strips will provide a minimum of 9 hits (5 Pixel and 4 end-cap).

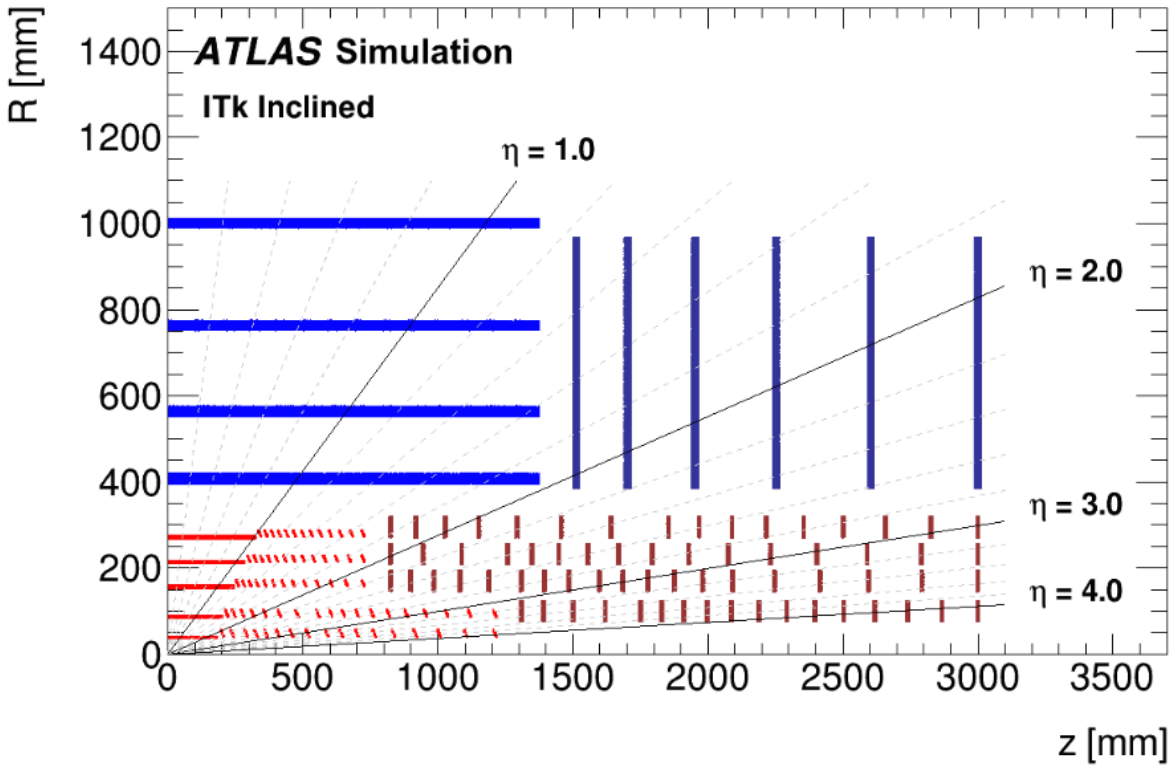


Figure 4.1: The ATLAS ITk layout [31].

The Strip upgrade consists of 11 types of modules for the two types of geometries barrel (2 modules) and end-cap (9 modules). They are made up from a few principle components apart from the silicon sensors themselves. These are :

- Hybrids - a low power PCB hosting the readout ASICs and controllers,
- ABCStar the readout electronics (ASICs).
- HCCStar hybrid controller chip,
- power-board.

The modules are summarised below in Table 4.1.

Module Name	Hybrid(s)	No. Power boards	Total no. ABCStar	Total no. HCCStar
SS	X, Y	1	20	2
LS	Y	1	10	1
R0	R0H0, R0H1	1	17	2
R1	R1H0, R1H1	1	21	2
R2	R2H0	1	12	2
R3S0	R3H0, R3H2	1	14	-
R3S1	R3H1, R3H3	-	14	4
R4S0	R4H0	1	8	-
R4S1	R4H1	-	8	2
R5S0	R5H0	1	9	-
R5S1	R5H1	-	9	2

Table 4.1: Table of ITk Strip modules [31].

The end-caps are made from 6 petals each made up from 9 modules as shown in Figure 4.2 and so have a curved geometry. Unlike the barrel modules not every module will have its own power board or controller chip.

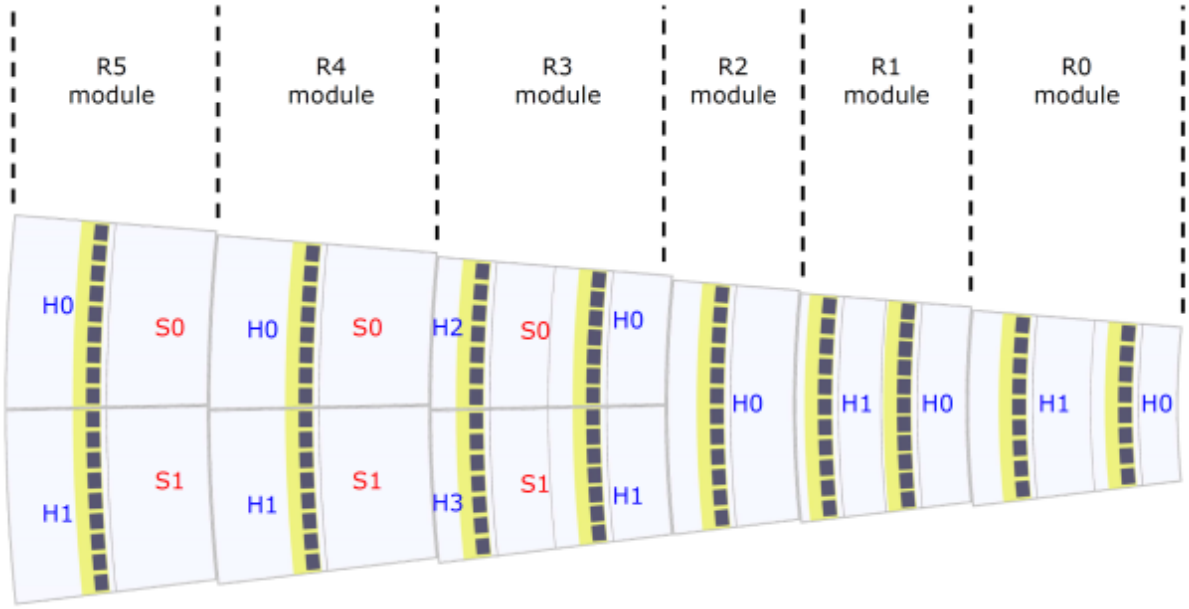


Figure 4.2: The ATLAS ITk end-cap petal layout [31].

Assembly of the modules requires several gluing steps to take place. The ASICs are glued onto the hybrids which are then glued onto the sensor. This needs to happen in such a way that allows for the readout electronics on the hybrids (the ASICs) to be wire bonded to the rest of the sensor. The power-boards are also glued onto the sensor but are less of an issue.

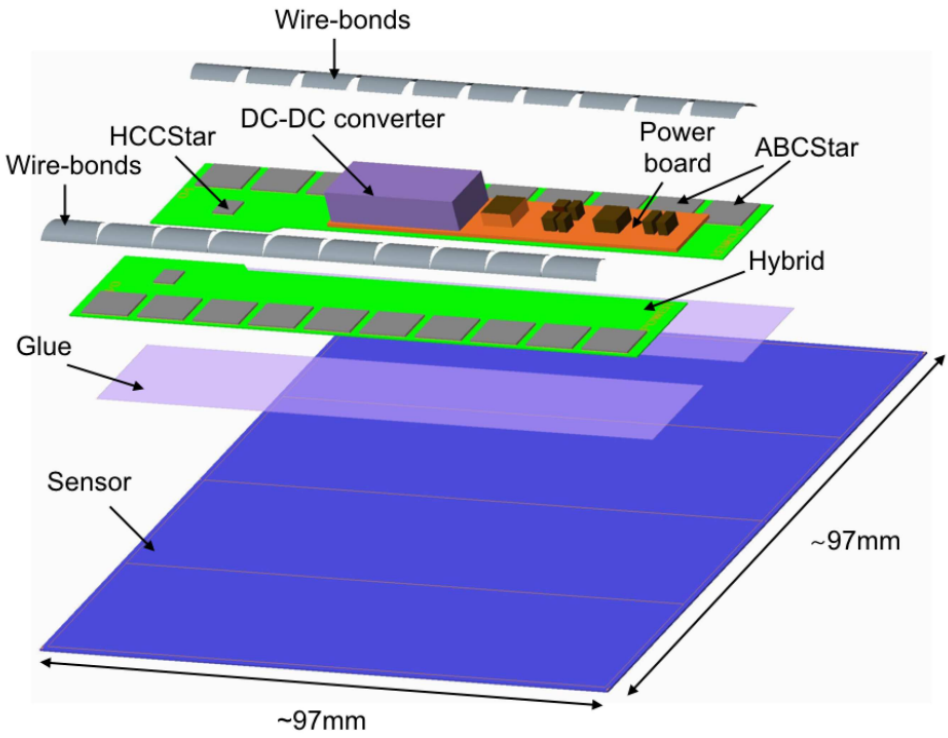


Figure 4.3: The ATLAS ITk SS module gluing of assembly step [31].

The wire bonding from ASIC to sensor can be tricky as tails can form during the process. These tails extend out in-line with the wire-bond and can short to other bond-pads or wire bonds (Figure 4.4). This puts a limit on the maximum wire-bond angle (22°) and puts a constraint on the maximum x-y error allowed during component placement. Module measurements will be taken to ensure wire-bondability.

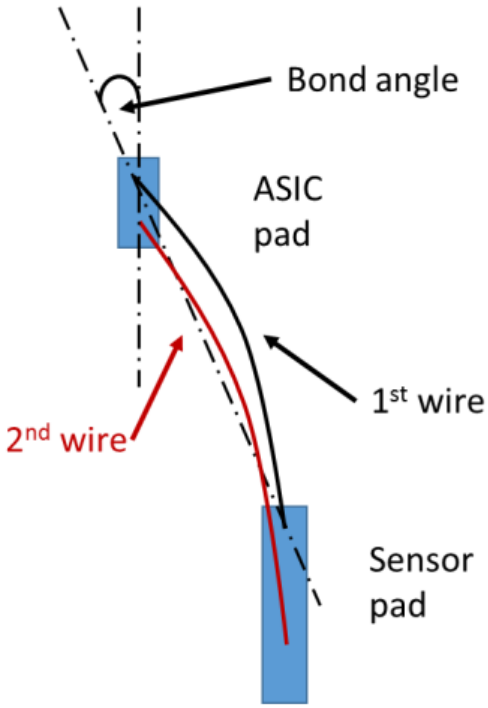


Figure 4.4: The bond angle with the vertical [31].

My qualification task description was to:

- Finalise a way that all the 22 module assembly sites can output (for all module types) a set of module metrology measurements for the production database.
- Collect and share information on what systems all the module building groups have.
- Calculate a set of envelopes allowed for x-y placement of hybrids and power-boards on modules.
- Automate the outputs of the metrology systems to the database.

The ≈ 22 module building sites will have to take metrology measurements and to ensure that all building sites have adequate machinery to take the required measurements; a list was gathered of all the metrology systems at the various sites. The building sites were split into barrel and end-cap with the UK effort focused on the barrel modules.

The main focus was then shifted to developing measurements for the x-y placement of the components. As well as rotations/translations in the x-y plane there could also be issues with the z placement, both in glue thickness and in variations of the thickness, but these were left to another study. Rather than calculate all the possible envelopes a software package was written that calculates the placement of the components and outputs a pass fail criteria. The output can then be uploaded to the production database as a JSON file.

The Short Strip (SS) module (see Figures 4.3, 4.5) was then used as a test case to set-up and validate the software. The modules are defined in configuration files that can

be loaded when performing the relevant module metrology. The bottom left corner was chosen as the coordinate system as shown in (Figure 4.5). The hybrid, ASICs, ASIC bond-pads and HCCs can all be defined in coordinates relative to the hybrid, which is then also defined in the global coordinate system of the sensor. This then allows for the creation of compound objects on which translations and rotations can be applied.

The measurements are taken of fiducials located on the hybrids and ASICs, these are then fed into the script which then calculates the transformation between the measured and ideal points. It calculates the rotation directly as sin and cos values to avoid rounding errors, with the translation being a simple subtraction operation. Then the module is constructed in memory and transformed through the measured transformation. It is then checked for a series of criteria: wire bonding angles, overhangs and collisions.

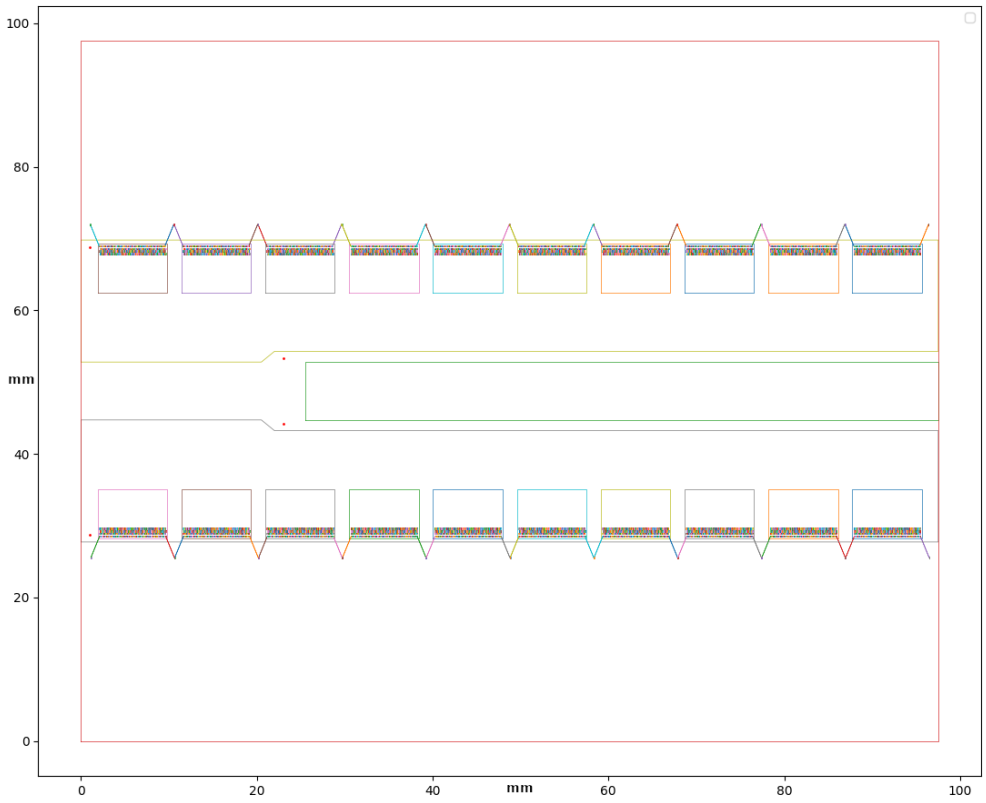


Figure 4.5: An SS module with ideal component placement, only the outermost wire bonds were modelled as these would have the most extreme angles, x-y axis are in mm.

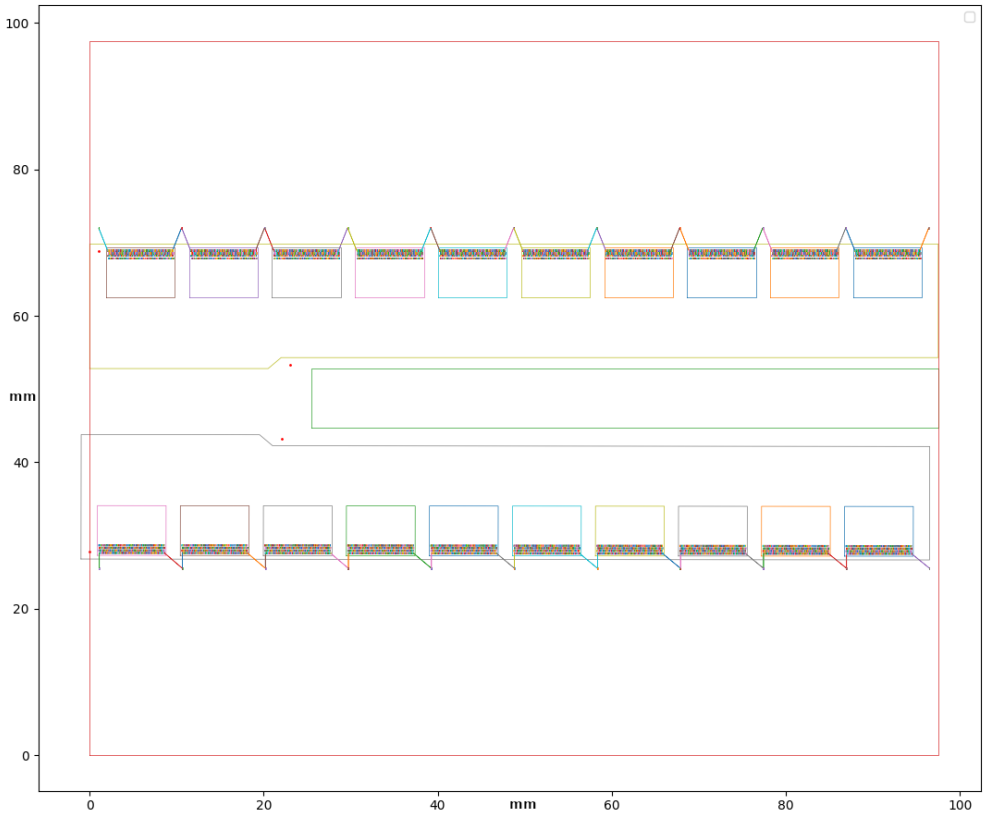


Figure 4.6: An SS module with the bottom hybrid placed 1 mm to the left.

A debug feature was added, a visualisation tool, used for testing and creating configuration files for modules. With it we can see the ideal placement of components for the SS module (Figure 4.5). If there is an error of 1 mm in the x-axis during assembly (Figure 4.6), then the module would not pass. This is because the bottom hybrid would overhang the sensor (Figure 4.6 at (0.0,43.80) and (0.0,26.78)) and in this case each ASICs right most wire bond-angle would be to large. For the ideal placement the min/max wire-bond angle is 18.43°/18.43° while in this case the min/max wire-bond angle is 5.73°/87.84°.

This information is then uploaded to the production database, used to keep track of the numerous parameters for each produced module, for both current manufacturing needs and for potential future reasons. The task was signed off on the bases of a report in the form of a readme for the software package developed; this can be seen in Appendix A.

Chapter 5

Monte-Carlo Simulation

In order to compare the data to theory Monte-Carlo (MC) integration is extensively used in simulating the processes involved. This is because of the theories inherently stochastic nature and is utilised throughout the simulation stages including the matrix element calculation and the detector simulation. The full simulation chain starts with the matrix element (ME) calculation which describes the hard scattering process of interest. The next step is to simulate the resulting soft interactions which, are far too numerous to calculate explicitly and is done with a parton shower (PS). After this the hadronisation of the resulting particles is simulated with a non-perturbative model. Following this the remnants of the proton can also enter the detector and will need to be accounted for in a process referred to as the underlying event (UE). On average there are 30-40 interactions present in one bunch crossing resulting in signals contaminating the hard scattering event of interest. These processes are referred to as pileup (PU). The final stage is to simulate the detector effects, with its specific geometry, and the various interactions particles undergo as they propagate through it. These stages are depicted in Figure 5.1.

5.1 Matrix Element Generation

Hard scattering events from the pp collision lead to the interesting processes that one looks for to test the SM. In order to make a prediction one must calculate the cross-section of the given process, which essentially is the probability that it will occur. In order to do this one typically uses perturbation theory and expands around the relevant coupling constant. However, when dealing with QCD interactions the strong coupling constant a_s decreases with energy meaning that when $a_s \to 1$ approaches unity an expansion will no longer converge and perturbation theory becomes unreliable. Because we are dealing with pp collisions they naturally contain a mix of high and low energy interactions resulting from the colliding partons and the internals of the protons respectively. To get around this one uses the factorisation theorem.

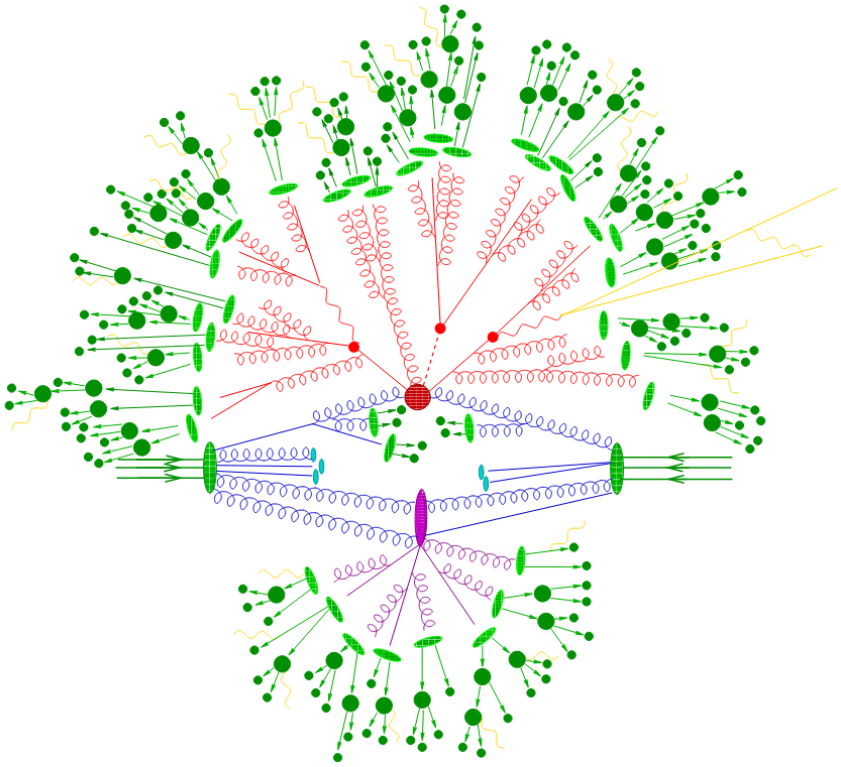


Figure 5.1: Depiction of a $t\bar{t}H$ event decaying semi-leptonically produced with an event generator. The incoming partons of the protons are represented by dark green blobs with three parallel arrows pointing towards the centre. The hard scattering event is represented by the big red blob and is induced via gluon-gluon fusion. The parton shower is shown in blue. The decay of the tops/Higgs are represented with small red blobs with the $t\bar{t}H \rightarrow bq\bar{q}, \bar{b}l\mu, b\bar{b}$ decay mode. Additional hard QCD radiation is shown in red and secondary interactions are shown in purple. The light green shows the particles prior to hadronisation and the dark green represents the hadrons decaying. Photon radiation is depicted in yellow [32].

5.1.1 Factorisation Theorem

The factorisation theorem [33] states that when calculating a cross-section one can define a factorisation scale (μ_f) at the boundary where perturbation theory can accurately describe the phenomena. This allows the the calculation to be factorised into high and low energy parts which can be solved independently. Given two hadrons with momentum P_1 and P_2 its cross-section can be factorised into:

$$\sigma(P_1, P_2) = \sum_{ij} \int dx_1 \; dx_2 \; f_i(x_1, \mu_f^2) \; f_j(x_2, \mu_f^2) \; \hat{\sigma}_{ij}(p_1, p_2, \alpha_s(\mu_f^2), Q^2/\mu_f^2).$$

Where $p_i = x_i P_i$ denotes the hard scattering parton's momentum fraction and Q denotes the interaction scale. Thus the perturbative piece only contains interactions with energies above μ_f^2 and is dependant on the coupling constant α_s. All the other interactions, with energies below μ_f^2, are gathered into the parton distribution functions (PDFs) $f(x, \mu_f^2)$,

which are summed over the indices i, j. Thus perturbation theory can be used to calculate the cross-section and decays of a process such as $t\bar{t}H$ due to the hard scattering partons, with the effects from the parton-parton interactions internal to the proton being described by the PDFs.

5.1.2 Parton Distribution Functions

The internals of a proton are described by Parton Distribution Functions (PDFs), consisting of three valance quarks (uud) and a quark-gluon sea. Each parton carries some fraction of the protons total energy. It is this distribution that the PDFs describe, with the DGLAP [33] equations describing the PDFs energy dependence. The momentum fraction of each parton (x) cannot be calculated analytically and is instead fit from experimental data. Deep inelastic scattering (DIS) experiments such as ZEUS and H1 [34, 35] use electron-proton collisions to probe the structure of protons and provide the experimental data for the PDF fits. These fits can then be further refined with the use of additional LHC experimental data.

5.1.3 Fixed Order Expansion

Typically the hard scattering processes of interest are calculated via a fixed order perturbative expansion in α_s. It must take into account both ultraviolet (UV) and infrared (IR) divergent effects. UV divergences are caused by high energy loops and IR divergences are caused by low energy collinear radiation. In order to address UV divergences, which can appear at any order, a renormalisation scale (μ_r) should be chosen taking into account the specific hard scattering process of interest [36]. As such a reasonable choice for both μ_r and μ_f, both of which are freely chosen, would be $\mu_r = \mu_f = Q^2$.

Expansions beyond leading order become increasingly complex while providing ever decreasing contributions. Once the magnitude of the contribution becomes comparable to the uncertainties in the measurement there is little to be gained by including them. This together with the fact that higher order corrections require more computational resources gives rise to a fixed order to expand to for a given measurement. Typically $t\bar{t}H$ predictions are available up to next to leading order (NLO) [37]. In order to create a more representative prediction a parton shower is used to simulate additional radiation.

5.2 Parton Shower

Once the hard process has been generated at a fixed order its evolution is modelled with a parton shower. It is used to simulate the higher order effects due to lower energy emissions. There is a probability of a parton spontaneously splitting into two, each carrying

momentum fractions z and $(z-1)$ with an angular separation of θ. The parton shower is calculated independently of the original hard scattering calculation using the outgoing particles final states and applying the relevant splitting probabilities. The parton splittings are ordered by momentum or angle and calculated with MC. It is controlled with the Sudakov Form Factor which describes the probability of a particle evolving from state t_0 to t:

$$\Delta(t_0, t) = \exp\left[-\int_t^{t_0} \frac{dt'}{t'} \int_{z_{min}}^{z_{max}} dz P(z)\right],$$

where $p(z)$ is the splitting function for quark/gluon splitting. In order to avoid IR and collinear divergences a cut-off energy scale must be defined typically around ~ 1 GeV. Care must be taken when matching the PS to the result of the fixed order expansion at NLO in order to avoid double counting. There are two general strategies used at NLO either: emissions from the PS are only taken below a certain threshold energy, ensuring that the additional higher emissions originate from the NLO ME expansion; or by finding the highest energy emissions in the PS and subtracting them from the NLO ME expansion, thus avoiding any overlap.

5.3 Hadronisation

After the PS has been applied the remaining particles will undergo hadronisation due to quark confinement with the resultant hadrons being colour neutral states. Due to the process being below the energy scale where a perturbative expansion is convergent it is approximated with models driven by experimental data. The two primary approaches are: the Lund string model [38] and the cluster model [39]. In the Lund string model the quarks are connected by a string of gluons with colour charge. As the quarks move further apart the potential energy of the string grows until it becomes energetically favourable for the string to break and pull a $q\bar{q}$ pair out of the vacuum, attaching to the newly formed ends of the string. Because of this the two newly formed strings are shorter and have lower energies, thus dissipating the overall energy. This process is repeated until the initial quarks no longer have the energy to stretch the strings to a breaking point. Alternatively the cluster model takes the gluons left over from the parton-shower to decay into $q\bar{q}$ pairs, which together with the other final state quarks form colour singlet clusters due to preconfinement. These clusters will then further decay until stable hadrons are produced.

5.4 Underlying Event / Multi-parton Interactions

The remaining partons from the initial protons, not involved in the hard scattering process, also undergo secondary interactions which will create additional particles that will be detected and is known as the Underlying Event (EV). Additional interactions between remnant partons are referred to as Multi-parton Interactions (MPI). The remnant particles will also undergo hadronisation. To account for these effects in the MC simulation dedicated MPI models are used with their parameters being tuned to experimental data.

5.5 Detector Simulation

The final step in recreating the final stage events as seen by the detector is simulating how the final state hadrons / particles (e.g. electrons, photons and stable hadrons) propagate through the detector so as to make a direct comparison with data. As these particles propagate through the detector they undergo various interactions with the detector subsystems resulting in processes such as hadronic showers, EM showers from electrons/photons etc. In order to do this, precise modelling of the detector geometry is required to account for the known resolutions of the detector. Typically the GEANT 4 [40] framework is used for ATLAS detector simulation. This is referred to as Full Simulation (FS) and uses MC methods to simulate individual particles as they propagate through the detector. While this method is rather accurate it is extremely CPU intensive with it requiring the majority of ATLAS' CPU time, especially for analyses that require large statistics and by extension involve a large number of particles. It is the simulation of the ATLAS calorimeter that is particularly CPU intensive, to get around this a dedicated tool called Atlfast-2 (AF2) [41] was developed by ATLAS where the calorimeter description is parametrised, with the FastCaloSim package [42], which drastically reduces the CPU overhead by an order of magnitude. However, this is accompanied by a somewhat poorer description of the calorimeter response and thus poorer agreement with data. This effect can be mitigated to some degree by dedicated calibrations resulting in an overall good agreement with data for the high level objects of interest.

Chapter 6

Analysis Techniques

Many analyses share techniques and strategies that are commonly used in LHC analyses. For the analyses described in Chapter 8 the signal is measured with a profile likelihood fit that is described in Section 6.1. Machine learning algorithms are used in many different stages of the analysis from simulation to reconstruction. Two common models used in particle physics are Boosted Decision Trees (BDTs) and Deep Neural Networks (DNNs), they are commonly used as particle and event classifiers but can also aide in other aspects such as detector simulation. These algorithms are discussed in Section 6.2.

6.1 Statistical Methods

When searching for a new process one can have two outcomes, signal discovery or limit setting. When trying to discover a new process the null hypothesis (H_0) is regarded as all currently known (background) processes and the alternative hypothesis (H_1) is regarded as the combination of the background and signal processes. However, when setting constraints H_0 becomes the signal plus background and the hypothesis H_1 becomes the background only hypothesis.

To quantify the level of agreement between the observed data and the hypothesis H one needs to calculate the p-value which, gives the probability that one can find data of equal or grater incompatibility than is predicted by H. To determine the level of incompatibility a likelihood ratio test may be used.

The reason for the 5σ convention is that in that past 3σ "discoveries" have turned out to be statistical fluctuations due to the fact that if you look at an ever increasing data set then the number of possibilities also increases. This is called the look elsewhere effect (LEE) [43] and can have up to a 100 fold boost in probability depending on the channel. With this effect in mind the 5σ convention was set as it was deemed rare enough that this would not be a problem and corresponds to a probability (p-value) of 0.0000003 (1 in 3.5 million). Here σ is used to refer to standard deviations of the normal distribution and can

be used to define statistical significance (Z) as in equation 6.14.

6.1.1 Preliminaries

Take an experiment where some kinematic variable x is to be measured such that the data is represented in a histogram $\vec{n} = (n_1, \ldots, n_N)$, with N bins and n_i being the number of entries in each bin. Then the expectation value of the n_ith bin is

$$E[n_i] = \mu s_i + b_i, \tag{6.1}$$

where s_i and b_i are the mean number of entries in each bin such that:

$$s_i = s_{\text{tot.}} \int_{\text{bin}_i} f_s\left(x; \vec{\theta} s\right) dx, \tag{6.2}$$

$$b_i = b_{\text{tot.}} \int_{\text{bin}_i} f_b\left(x; \vec{\theta}_b\right) dx. \tag{6.3}$$

Here $s_{\text{tot.}}$ / $b_{\text{tot.}}$ are the total mean number of events for signal/background and f_s / f_b are the probability density functions (PDFs) for signal/background which are dependent on the variable x and are characterised by parameters $\vec{\theta}_s$ and $\vec{\theta}_b$ respectively. The μ parameter above represents the strength of the signal model *i*.e $\mu = 0$ implies a background only hypothesis and $\mu = 1$ implies the signal plus background hypothesis.

Nuisance parameters are parameters whose values are not known *a priori* and must be fitted from data. These terms are gathered together as $\vec{\theta} = \left(\vec{\theta}_s, \vec{\theta}_b, b_{tot}\right)$. It should be noted that $s_{\text{tot.}}$, the signal normalisation, is not a nuisance parameter but is predicted from the signal plus background hypothesis.

In order to constrain the nuisance parameters one often makes a control measurement in a region where the background process dominates. This can give information on the background normalisation parameter $b_{\text{tot.}}$ and potentially some of the other nuisance parameters as well. This is denoted as histogram $\vec{m} = (m_1, \ldots, m_M)$ with an M number of bins and expectation value

$$E[m_i] = u_i(\vec{\theta}), \tag{6.4}$$

where u_i refers to a specific background process and depends upon the nuisance parameters $\vec{\theta}$. Typically one would construct the control measurement to fit the background normalisation (b_{tot}) and shape ($\vec{\theta}_b$) parameters.

6.1.2 Likelihood ratio test

The likelihood function of the above mentioned measurements (for both N and M) is a product of the Poisson probabilities:

$$L(\mu, \vec{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}, \tag{6.5}$$

which gives the likelihood of observing the expectation values $E[n_i]$ and $E[m_k]$ given n_i and m_k events.

In the testing of a hypothesised value of μ one can employ the likelihood ratio test:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\vec{\theta}}})}{L(\hat{\mu}, \hat{\hat{\vec{\theta}}})}, \tag{6.6}$$

where the parameters $\hat{\hat{\vec{\theta}}}$ are chosen such that they maximises L for a given μ. The parameters $\hat{\mu}$, $\hat{\vec{\theta}}$ are chosen to maximise L. So the numerator is a conditional (on μ) maximum likelihood (ML) estimator and the denominator is an unconditional maximum likelihood (ML) estimator.

From Wilk's theorem [44] this can be related to a *chi-squared* distribution (χ_0^2) allowing for the calculation of a p-value through numerical methods. The relation is:

$$-2 \ln \lambda \approx \chi_0^2, \tag{6.7}$$

with one degree of freedom.

6.1.3 Distribution of the likelihood ratio test

When testing a hypothesised value for μ it is convenient to use a test statistic $t_\mu = -2 \ln \lambda(\mu)$. So the lower the value of t_μ the more agreement there is between data and hypothesis. The p-value for an observed value of $t_{\mu,\text{obs}}$ then is

$$p_\mu = \int_{t_{\mu,\text{obs}}}^{\infty} f\left(t_\mu | \mu\right) dt_\mu, \tag{6.8}$$

with $f\left(t_\mu | \mu\right)$ being the PDF of t_μ.

A low p-value may be obtained if the estimated signal strength $\hat{\mu}$ is found to be greater or less than the hypothesised value of μ. From this one can reject values of μ, given a required p-value, resulting in a two sided confidence interval.

When looking at a test for a hypothesised value of μ (from equation (6.1)) and given that the data is distributed according to μ' the distribution $f(t_\mu | \mu')$ can be approximated by the Wald approximation [45]:

$$-2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}). \tag{6.9}$$

This holds in the limit of a large sample size (e.g. $N \to \infty$). To use equation (6.9)

one needs to calculate the standard deviation σ of the Gaussian distribution $\hat{\mu}$ with mean μ'. From this it can be shown that $t_\mu = -2\ln\lambda(\mu)$ follows a non-central chi-squared distribution for one degree of freedom:

$$f\left(t_\mu;\Lambda\right) = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}+\sqrt{\Lambda}\right)^2\right)+\exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}-\sqrt{\Lambda}\right)^2\right)\right], \quad (6.10)$$

with the non-centrality parameter λ being

$$\Lambda = \frac{\left(\mu-\mu'\right)^2}{\sigma^2}. \quad (6.11)$$

In the case of $\mu' = \mu$ one has $\Lambda = 0$ and $-2\ln\lambda \approx \chi^2_0$ as shown by Wilks [44].

Generalising to more than one parameter of interest ($\vec{\theta_r} = (\theta_1, \ldots, \theta_r)$) the distribution $-2\ln\lambda(\vec{\theta_r})$ then approaches a non-central chi-squared distribution with r-degrees of freedom. The non non-centrality parameter then becomes

$$\Lambda_r = \sum_{i,j=1}^{r}\left(\theta_i-\theta_i'\right)\tilde{V}_{ij}^{-1}\left(\theta_j-\theta_j'\right), \quad (6.12)$$

where $\tilde{V}_{ij}^{-1}$ is the inverse of the sub-matrix of the full covariance matrix. The covariance matrix ($V_{ij} = \mathrm{cov}[\hat{\theta}_i, \hat{\theta}_j]$) of the ML estimators is the standard deviations of the parameters (both of interest and nuisance). It's inverse is given by

$$V_{ij}^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial\theta_i\partial\theta_j}\right], \quad (6.13)$$

which holds in the large sample limit.

6.1.4 Standard deviation and significance

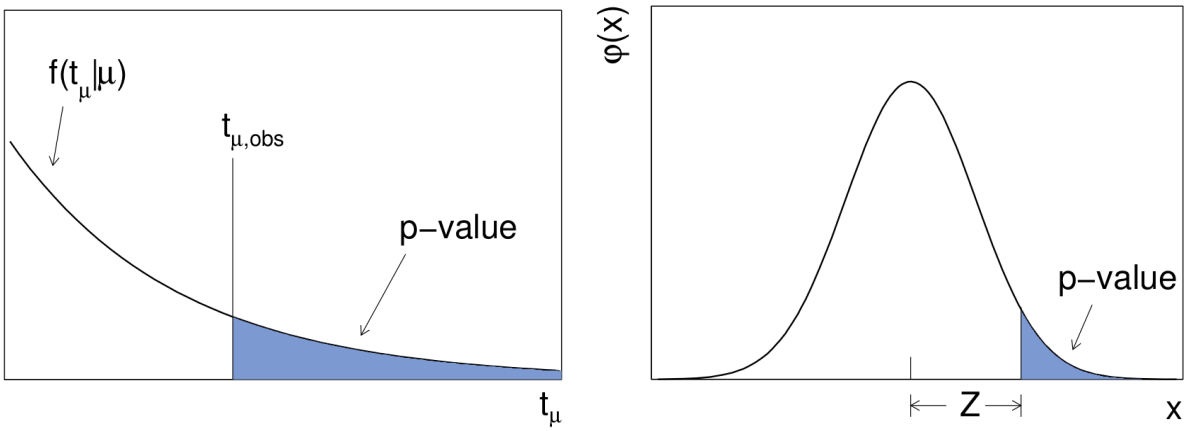
The p-value for a test statistic t_μ is the area under its curve, as given by equation (6.8), from this a significance of Z standard deviations can be obtained with,

$$Z = \Phi^{-1}(1-p), \quad (6.14)$$

where Φ^{-1} is the quantile of the standard Gaussian. This is illustrated in Figure 6.1 below. Quantiles are divisions of a probability distribution into equally probable sections.

6.1.5 Asimov data set

The Asimov data set is an artificial data set which can be very useful in estimating $\sigma_{\hat{\mu}}$ (where $\hat{\mu}$ is assumed to be a Gaussian distribution with mean μ'). It is defined such that



(a) (b)

Figure 6.1: Shows the p-value (shaded) for the observed test statistic $t_{\mu,\mathrm{obs}}$ on the left (a) and its equivalent Gaussian significance for the standard normal distribution $\varphi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$ on the right (b) [46].

each of the bins in its histogram is its expectation value. In other words it would be the "perfect experiment" with no statistical variation, for a simple coin flipping experiment of 100 flips the histogram of the Asimov data set would consist of two bins each with the expected value of 50 in them.

More formally let $\nu_i = \mu' s_i + b_i$ and $\theta_0 = \mu$, the strength parameter, so θ_i encompasses all the parameters. Then the ML estimators can be found by setting the derivatives of $\ln L$ given in equation (6.5) to zero:

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{N} \left(\frac{n_i}{\nu_i} - 1 \right) \frac{\partial \nu_i}{\partial \theta_j} + \sum_{i=1}^{M} \left(\frac{m_i}{u_i} - 1 \right) \frac{\partial u_i}{\partial \theta_j} = 0. \tag{6.15}$$

This will be true because the Asimov data is equal to its expectation values:

$$n_{i,\mathrm{A}} = E[n_i] = \nu_i = \mu' s_i(\vec{\theta}) + b_i(\vec{\theta}), \tag{6.16}$$

$$m_{i,\mathrm{A}} = E[m_i] = u_i(\vec{\theta}). \tag{6.17}$$

From results shown in [46] the Asimov likelihood is approximately

$$-2\ln \lambda_{\mathrm{A}}(\mu) \approx \frac{(\mu - \mu')^2}{\sigma^2} = \Lambda, \tag{6.18}$$

which is the non-centrality parameter that characterises the PDF, thus giving a relation between the Asimov likelihood and the variance (σ^2).

6.1.6 Test statistics

When looking at the test statistic t_μ and the likelihood ratio test $\lambda(\mu)$ in equation (6.6) it can be seen that the ML estimator can take on negative values ($\hat{\mu} \leq 0$). However, when looking for a new signal one often assumes that the new signal will increase the event rate over the already known background. To this effect $\tilde{t}_\mu$ is defined below with $\mu \geq 0$ from its corresponding likelihood ratio $\tilde{\lambda}(\mu)$

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\vec{\theta}})} & \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{L(0, \hat{\hat{\vec{\theta}}}(0))} & \hat{\mu} < 0, \end{cases} \tag{6.19}$$

where $\hat{\hat{\vec{\theta}}}$ is the conditional ML estimator given a strength parameter μ or 0. The corresponding test statistic $\tilde{t}_\mu$ is

$$\tilde{t}_\mu = -2 \ln \tilde{\lambda}(\mu) = \begin{cases} -2 \ln \frac{L(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\vec{\theta}})} & \hat{\mu} \geq 0, \\ -2 \ln \frac{L(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{L(0, \hat{\hat{\vec{\theta}}}(0))} & \hat{\mu} < 0. \end{cases} \tag{6.20}$$

A specific example of the above mentioned test statistic $\tilde{t}_\mu$ is used for the discovery of a positive signal. In this case one is testing for $\mu = 0$ when it is assumed that $\mu \geq 0$, if one can reject the $\mu = 0$ hypothesis then it can be said a new signal has been discovered. This special case of $\mu = 0$ will be denoted by $q_0 = \tilde{t}_\mu$ and is

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0. \end{cases} \tag{6.21}$$

When using this test statistic if $\hat{\mu} > 0$ then it is to be read the same as t_μ, however, if $\hat{\mu} < 0$ then it is inconclusive and while it may mean a disagreement with the background only hypothesis it may also be from another systematic error.

When setting upper limits one could employ a similar test statistic

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu, \end{cases} \tag{6.22}$$

where when $\hat{\mu} > \mu$ it is inconclusive.

It should be noted that as the test statistics q_μ and q_0 are testing for different things their inconclusive regions correspond to different things. For q_μ ($\hat{\mu} > \mu$) corresponds to an upward fluctuation in data but for q_0 ($\mu < 0$) corresponds to a downward fluctuation in data. In both cases p-values and thus significances can be calculated however, to do this

one must be able to approximate the distribution of the PDFs, as described in [46] and gives

$$Z_0 = \sqrt{q_0}, \tag{6.23}$$

and

$$Z_\mu = \sqrt{q_\mu}. \tag{6.24}$$

6.1.7 Sensitivity of an experiment

When determining the sensitivity of an experiment one is not interested in the significance of a specific data set or MC sample but rather would like to determine what is the expected significance given an unlimited access to new data sets. To this end one wants to obtain the median values of q_0 and q_μ for which the Asimov data set is used.

$$\text{med}\left[Z_0|\mu'\right] = \sqrt{q_{0,\text{A}}}, \tag{6.25}$$

$$\text{med}\left[Z_\mu|0\right] = \sqrt{q_{\mu,\text{A}}}, \tag{6.26}$$

where $\text{med}\left[Z_0|\mu'\right]$ and $\text{med}\left[Z_\mu|0\right]$ are the median significance values from the test statistics $q_{0,\text{A}}$ and $q_{\mu,\text{A}}$ calculated from the Asimov data set.

It is also useful to be able to combine multiple channels so as to achieve the maximum possible sensitivity for a given process (e.g. $t\bar{t}H$). Given that for each channel i there is a likelihood function $L_i\left(\mu, \vec{\theta}_i\right)$, where $\vec{\theta}_i$ are the nuisance parameters for that channel. Then if the channels are statistically independent the channels likelihood function is given by

$$L(\mu, \vec{\theta}) = \prod_i L_i\left(\mu, \vec{\theta}_i\right). \tag{6.27}$$

So the profile likelihood ratio is

$$\lambda(\mu) = \frac{\prod_i L_i\left(\mu, \hat{\hat{\vec{\theta}}}_i\right)}{\prod_i L_i\left(\hat{\mu}, \hat{\hat{\vec{\theta}}}_i\right)}, \tag{6.28}$$

since there are no statistical fluctuations in the Asimov data set $\hat{\mu} = \mu'$ for all the channels. This combined with the fact that if channels share nuisance parameters then those parameters are equal in all the channels, one finds the profile likelihood ratio for all channels given a strength parameter μ' to be

$$\lambda_\text{A}(\mu) = \frac{\prod_i L_i(\mu, \hat{\hat{\vec{\theta}}})}{\prod_i L_i\left(\mu', \vec{\theta}\right)} = \prod_i \lambda_{\text{A},i}(\mu). \tag{6.29}$$

When combining channels and finding the discovery significance or exclusion limits from real data one needs to construct the full likelihood function with a single parameter μ which can be used in a global fit to find the profile likelihood ratio.

6.1.8 Example of a counting experiment with a known background

Given an experiment with a known background that follows a Poisson distribution with an expectation value for n events of $E[n] = \mu s + b$. Then since we can treat b as a constant to obtain the likelihood function of

$$L(\mu) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)}, \tag{6.30}$$

so the test statistic in the case of a discovery is

$$q_0 = \begin{cases} -2 \ln \frac{L(0)}{L(\hat{\mu})} & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases} \tag{6.31}$$

with $\hat{\mu} = n - b$. Assuming the asymptotic approximation of Feldman and Cousins [47] then the significance is given by equation (6.23)

$$Z_0 = \sqrt{q_0} = \begin{cases} \sqrt{2\left(n \ln \frac{n}{b} + b - n\right)} & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0. \end{cases} \tag{6.32}$$

To then get the median significance for the nominal signal hypothesis of $\mu = 1$ one can use the Asimov value for n of $s + b$ to obtain

$$\text{med}\left[Z_0|1\right] = \sqrt{q_{0,A}} = \sqrt{2((s+b)\ln(1+s/b) - s).} \tag{6.33}$$

This can be further reduced by expanding the logarithm in series around s/b to give

$$\text{med}\left[Z_0|1\right] = \frac{s}{\sqrt{b}}(1 + \mathcal{O}(s/b)). \tag{6.34}$$

This approximation is widely used and is valid for $s \ll b$.

6.2 Machine Learning Techniques

Machine Learning is a branch of computer science that has its roots in statistics [48–50]. It aims to create an algorithm that learns from its mistakes by minimizing an error metric. There are three main categories: supervised, unsupervised and reinforcement learning.

Supervised learning is where the data is labelled, unsupervised learning makes use of unlabelled data and reinforcement learning is a mixture of the two. Within supervised learning there are two classes of problems: classification and regression. They both work on the basis of fitting a function to the training data set, with classification using the function to separate the classes whereas regression predicts unknown values of said function.

The two ML algorithms that will be focused on here are Boosted Decision Trees (BDTs) and Deep Neural Networks (DNNs) with both of them being used in the classification context. This can be either as a tagger to "tag" a jet i.e b-jets as coming from a specific decay or as an event classifier where a specific event is being determined to have originated from a specific hard scattering process. Typically event classifiers make use of information from taggers.

6.2.1 Boosted Decision Trees

Boosted Decision Trees (BDTs) build upon the decision tree concept using so-called "weak learners". A decision tree makes use of the input features to separate the data and learns by minimising a loss function. Each feature will have a corresponding level where the tree applies a cut. The number of nodes in a level depends on the depth and branching, each node will have its own cut value. It is through the adjustment of these cut values that a decision tree learns. By themselves decision trees are considered "weak learners" meaning that they only perform marginally better than pure chance. To improve their performance an ensemble of trees can be used with two common strategies being random forests and boosting. Boosting combines N trees and iteratively modifies the training data set to ensure that the trees are not identical. Specifically successive trees are trained with ambiguous events, as determined by the previous tree, being weighted up. An illustration of this event weighting can be seen in Figure 6.2, with the final ensemble of trees being combined into a "strong learner". Two common boosting algorithms are AdaBoost [51] and GradientBoost [52].

6.2.2 Deep Neural Networks

Deep Neural Networks (DNN) are neural networks (NN) with several layers (at least 3). Typically there are 3 parts, the input layer, middle (hidden) layers and output layer. Input features are fed into the first layer, as such their dimensions must match. Then the hidden layers try to approximate the outputs of the function. This is done during training, over several cycles of the data, by adjusting the weights of the activation functions of the neurons. The aim of the algorithm is to minimize a loss function and with each new cycle the loss function is compared to see if progress has been made. Each neuron takes in the previous layers outputs as inputs. This is a weighted sum of all the outputs of

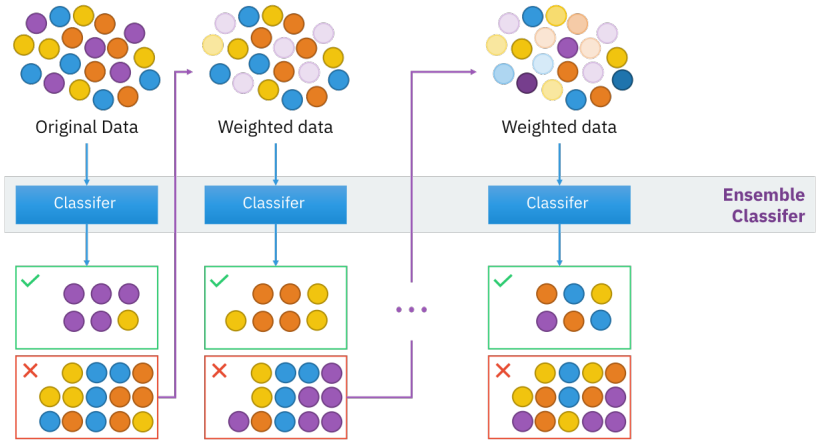


Figure 6.2: An illustration of the boosting algorithm successively weighting events and the ensemble of trees they create [53].

the previous layer. Each output has a specific weight it gets multiplied by, unique to that specific connection. It is through the adjustments of these weights that a NN learns. Then it has some activation function to see if it has been given a high enough value to activate itself. A neuron will only output a value if it has been "activated".

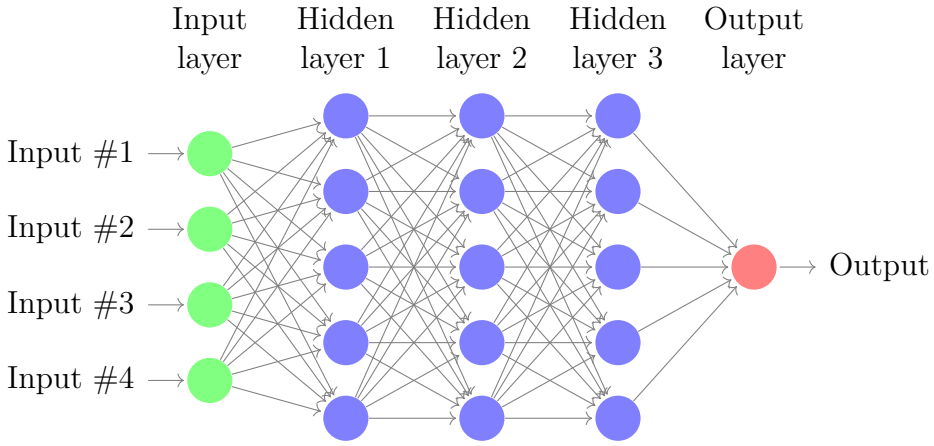


Figure 6.3: An example of a Deep Neural Network.

In the case of a simple Neural Network (NN) as in Figure 6.4, it can be seen that for each node (h_i^1) in the hidden layer $h_i^1 = a(\sum_j x_j w_{ji}^1 + b_i^1)$, where a is the activation function, x_j is an element of the input vector, w_{ji}^1 is an element of the first weight matrix (one for each connection) and b_i^1 is the bias for that node. Here the upper indices are used to denote the layer of the network for matrix elements. So the output vector (Y) is $((XW_1 + B_1)aW_2 + B_2)a = Y$, in this example W_1 is a 2x3 matrix, W_2 is a 3x1 matrix, B_1 is 1x3 matrix and B_2 is a 1x1 matrix.

As previously stated during training a network learns through the adjustment of the elements of these weight and bias matrices, in this case there are a total of 13 elements

for the network to play with. Although this is a fairly simple network it can easily be seen that the more complex a network gets the more elements the network has to adjust and the more computational power it requires to do so. However, once trained a NN simply has to multiply the input vector X by W a combined matrix of the weights matrices $W = a^n(W_1 + B_1)(W_2 + B_2)\dots(W_n + B_n)$. Here it is a 2x1 matrix, to make predictions and so is computationally inexpensive.

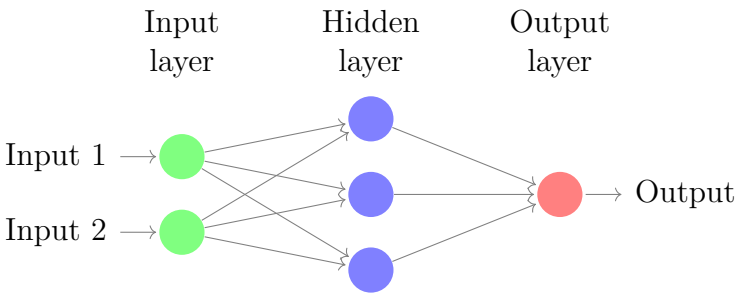


Figure 6.4: A simple Neural Network.

When creating a NN a number of parameters need to be defined, these are called the hyper-parameters. Although there is no clear way to chose these and typically trial and error is used, some insight can be gained. These parameters are as follows: the activation function, the optimization algorithm, the loss function, the initialization of the weights, the number of iterations it is trained for, the batch size of the input data and the width and the depth of the network.

Activation functions

In general activation functions are functions which activate when given an input large enough. Three of the most common activation functions are hyperbolic tangent (tanh), sigmoid (S) and ReLU (R), where $\tanh(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$, $S(x) = \frac{1}{1+e^{-x}}$ and $R(x) = \max(0, x)$ respectively. Up until recently the tanh and sigmoid activation functions were the most common however recently ReLU has become the standard. In practice ReLU can provide precision equal to that of the others provided the network is deep enough even when modelling non-linear processes. However, it is much faster to train [54] especially on deeper networks. If one is simply looking at how many operations are required to compute an activation function, then it can be seen that ReLU requires less than the others.

The ReLU activation function helps a NN obtain sparse (some 0 values in the matrix) representations of the data [55] which can be favourable because it helps isolate the parameters responsible for the variations in the data. Another advantage of ReLU is that it has a non-saturated gradient (has no upper limit on the image of the function) compared to the other functions which can saturate (the upper limit of the image being 1). When a node is saturated it becomes very difficult for that node to learn and consequently learns

very slowly [56]. Looking at the simple case in Figure 6.4 if a node was to become saturated then $h_i^k = a(\sum_j x_j w_{ji}^k + b_i^k)$ would have a value close 1. So any subsequent change in the weights w_{ji}^k or bias b_i^k would result in a very small change in h_i^k as the gradient of the functions in this region would be very small. A potential disadvantage may be that when $x < 0$ the activation function will also be 0, then these nodes would not be able to learn as any back-propagation of a gradient would not be possible subsequently hurting training. Experimental evidence seems to contradict this [55] with their hypothesis being that as long the gradient is able to back-propagate through some path the network can still learn.

The Softmax activation function is often used in the final layer of a classification DNN. It ensures that the total output probabilities sum to 1 i.e for a 3 class problem $P_1 + P_2 + P_3 = 1$. It is defined as:

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}},$$

and applies a normalisation by taking the exponential of each input element s_i and then divides them with the sum of the exponentials. When using this in conjunction with the Cross-Entropy loss function it will provide an approximately linear loss function in s_i allowing it to learn quickly and avoid a vanishing gradient.

Optimization

Back-propagation [57] is the process off updating the weights after each iteration. Initially a batch of input data is fed into the NN and is propagated through the network to calculate an output $Y = XW$ keeping each individual weight matrix W_k (if it's the first batch then these weights and biases will have to be initialized to some values), where X is the input batch, W is the combined weight matrix and Y is the output batch. Then the output Y is compared to the correct values $\hat{Y}$ with a loss function L and an error is calculated. These errors are then propagated back through the network so that each node has an associated error value e_i^k, where i is the index for nodes in a layer and k is the index of layers. In the case of Figure 6.4 the error is first fed into the $W_2 + B_2$ then the $W_1 + B_1$ matrix and would get error matrices E_2 (1x1) and E_1 (1x3). Once the error matrices have been computed the weights and biases can be adjusted through gradient descent, in order to minimize said loss function.

In general, gradient descent is the minimization of some function through several iterations. The derivative is taken at a point then a step is taken in the direction of the negative gradient (down) and the process is repeated at the new point. So the general algorithm for back-propagation for a network is:

1. Input a batch X, let $X = H_0$.

2. Feed the batch forward calculating $Z_k = H_{k-1}W_k + B_k$ and $H_k = aZ_k$.

3. Calculate the error $E_n = \nabla_H L \odot a'(Z_n)$.

4. Back-propagate the error $E_k = (E_{k+1}(W_{k+1})^T) \odot a' Z_k$

5. Update weights and biases by $w_i^k \to w_i^k - \frac{\eta}{m}\sum_x e_x^k(h_x^{k-1})$ and $b_i^k \to b_i^k - \frac{\eta}{m}\sum_x e_x^k$

Here $\odot$ is the Hadamard product,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \odot \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{pmatrix},$$

k is the layer index and n is the total number of layers, M^T is the transpose of M, a' is the derivative of a, η is the learning rate and m is the batch size. The learning rate determines how big a step is taken and is an important consideration because if it's too small then learning will be unnecessarily slow but too large then the minimum may be over stepped. Stochastic gradient descent is when some noise is added to the gradient which can jerk the model out of a local minimum. A momentum term can also be added which, combined with the gradient updates the weights, has the effect of helping the model not get stuck in steep ravines and flat plateaus.

Of course, this is only a simple case and many more variations exist, more recently the Adaptive Momentum Estimation (Adam) optimization algorithm [58] has enjoyed great success. It combines the benefits of two other extensions of stochastic gradient descent Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). AdaGrad's main advantage is that it maintains a per-parameter learning rate which helps with sparse gradients. RMSProp also maintains a per-parameter learning rate which is adapted by the average variance in the steepness.

Initialization, iterations and batch sizes

When initializing the weights and biases of a network typically they are set randomly [56] although it is important to make sure that they are within certain bounds dependent on the activation function. For example, if ReLU is being used a whole layer should never be set to 0 as this would block all back propagation and not allow the network to learn. Typically, its best to set the bounds within the 'active' region of the activation functions as this allows for quick learning. The 'active' region for a function would be a region where $|\frac{df}{dx}| > 0$ and is relatively large.

As it is generally not possible to pass the entire data set into the network at once we pass small batches of data to the network for it to learn from. The smaller the batch size the longer training will take, and the more updates will be done often gaining a factor of 2 in training time [59]. However, with larger batch sizes it is the convergence rate

that decreases [60] so a general sweet spot between training time and convergence time is needed to be found.

The number of iterations i depends on both the batch size b and the size of the data set d. Typically iterations $i \geq \frac{b}{d}$ as the network should at least be trained over all the data points. However, training over the same point repeatedly can lead to over fitting. Over fitting is when a network has learned a data set too well and no longer represents the general problem. This is a problem that can arise for several reasons one of them being the number of iterations being too high so picking a value of $i = \frac{b}{d}$ or only marginally greater is a safe bet. Another way to help prevent over-fitting is by introducing a dropout layer [61] into the middle of the NN. A dropout layer is one where for one iteration nodes and their weights are essentially ignored (dropped) at random for that iteration and the weight update is done without them. Those nodes and weights are then reinserted for the next iteration.

Width and depth

A NN has both a width and a depth, where the width corresponds to the number of nodes (neurons) in a layer and the depth corresponds to the number of hidden layers. A wider NN will have better memorization capabilities [62] and a deeper NN will have better generalization capabilities. A wide and deep NN will have properties of both but runs the risk of having so many parameters that it can simply learn the test set over-fitting to it and not generalize well. A dropout layer can also be used in this instance. Typically the larger the NN the larger the data set needs to be in order to train it effectively.

Implementation

The DNNs trained for Higgs reconstruction discussed in Chapter 8 were implemented and trained using the Keras/Tensorflow (developed by Google) [63, 64] python packages and then converted to C++ using the LwTNN [65] package for use within the analysis. Another common machine learning framework is PyTorch (developed by Facebook) and ONNX (developed by Facebook/Microsoft) has recently been gaining ground for DNN inference and can be used to convert a multitude of trained DNNs to C++ for fast inference and supports custom layers which is not readily available with LwTNN.

6.2.3 Loss Functions

Choosing a loss function is highly problem dependent [66,67]. Often the same loss functions can be used by both BDTs and DNNs. Typically loss functions for regression are mean squared error (MSE) $L = \frac{1}{n}\sum_i^n (y_i - \hat{y}_i)^2$ and mean absolute error (MAE) $L = \frac{1}{n}\sum_i^n |y_i - \hat{y}_i|$, where y_i is the predicted output and $\hat{y}_i$ is the true value.

For classification problems Cross-Entropy is commonly used. For multi-class problems implemented with DNNs Categorical Cross-Entropy is used and is combined with a Softmax (activation function) final layer. Cross-Entropy (CE) is defined as:

$$CE = -\sum_{i}^{C} t_i \log\left(f(s)_i\right).$$

6.2.4 Training

Both DNNs and BDTs are sensitive to over fitting whereby an algorithm (DNN/BDT) learns the training data set too well, no longer representing the general problem. There are various strategies to avoid this typically relying on either randomly varying training data sets (known as bagging for BDTs) or by randomly dropping certain connections (known as a dropout layer for DNNs). Stopping training at some "optimal" point before the algorithms become over trained whilst maximising their performance is the primary challenge in training ML algorithms. Cross validation is one method to gauge how well an algorithm is doing by splitting the data into a training and test set. When the performance of the algorithm starts to diverge between the training and test sets one can say that it's starting to over fit to the training set. k-fold cross validation takes this concept further by splitting the data in k folds. The algorithm can then be trained k times on each combination of $k-1$ folds. With the kth fold in each instance being used for testing as can be seen in Figure 6.5. This is particularly effective for hyper parameter optimisation of DNNs, once the optimal configuration has been found it can then be trained on the full data set for the final model. Another advantage of k-fold cross validation is that it can isolate specific regions to which the algorithms may be sensitive to over fitting which would show up as having significantly better performance in one instance and have a large standard deviation compared to the other runs.

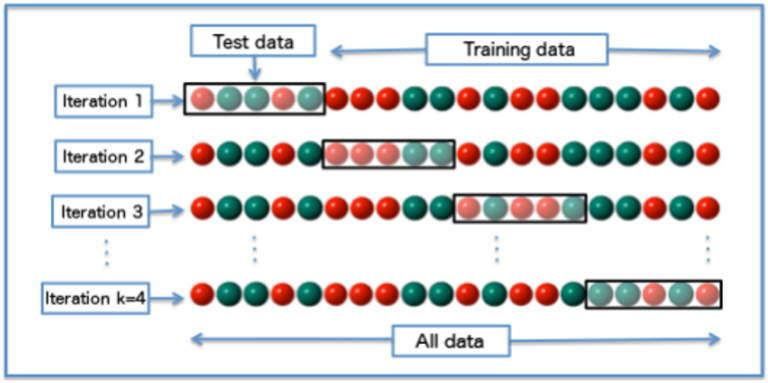


Figure 6.5: An example of k-fold cross validation with k set to 4 [68].

6.2.5 Hyper-parameter Optimisation

Optimising the hyper-parameters of a DNN can often lead to performance improvements. In order to do this some metrics evaluating the performance of the networks for comparison are needed. Cross validation should also be employed in order to gauge any over fitting and to help determine optimal stop conditions for the best results.

Metrics

The metrics used to evaluate the DNNs during training in the context of particle physics are:

- The specific loss function being used,
- $accuracy = \frac{TP}{Class+Other}$,
- $precision = \frac{TP}{TP+FP}$,
- $recall = \frac{TP}{Class}$.

Where TP is True Positive, FP is False Positive, $Class$ is the total number of true elements and $Other$ is the total number of false elements not belonging to the class. The combination of $TP + FP$ is the total number of elements classified and $Class + Other$ is the total number of elements seen. An example for these definitions can be seen in Figure 6.6.

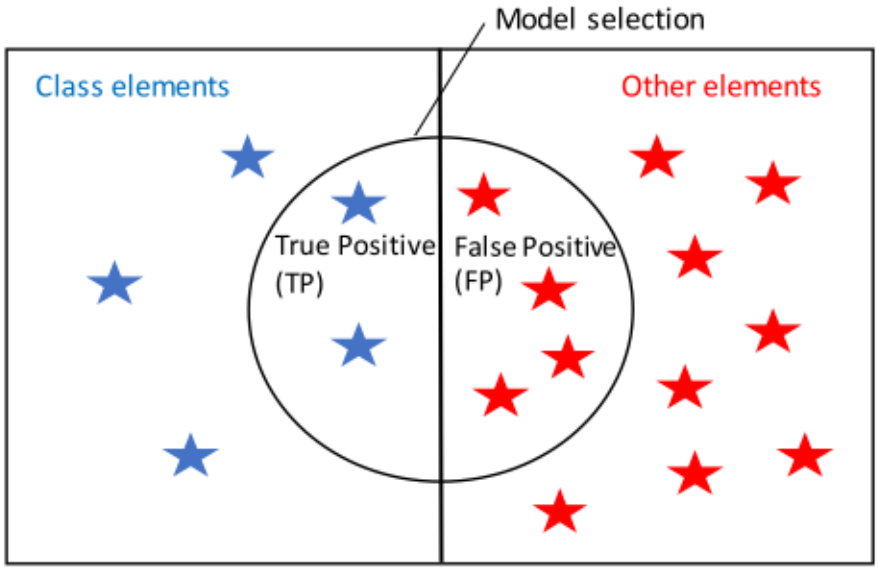


Figure 6.6: An illustration of true positive TP and false positive FP. $Class$ elements are indicated by blue stars and $Other$ elements are indicated by red stars.

Bayesian Optimisation

Bayesian Optimisation uses Bayes Theorem to direct the search in order to maximise the objective function, in this case the performance of our DNN. Bayes Theorem states that a conditional probability of an event is $P(A|B) = P(B|A) \times P(A)/P(B)$. Dropping the normalisation quantity $P(B)$ simplifies things but this then is no longer a probability, but a proportional quantity $P(A|B) = P(B|A) \times P(A)$, which we want to maximise/minimise. Typically the conditional probability is referred to as the *posterior probability*, the reverse conditional probability is referred to as the *likelihood* and the marginal probability as the prior giving: *posterior probability* = *likelihood* $\times$ *prior*. The prior is defined as a sequence of points x_n and their results $f(x_n)$ when evaluated by the objective function giving $D = \{x_0, f(x_0), ..., x_n, f(x_n)\}$. The likelihood function is defined as $P(D|f)$ and will continually be updated as new points are selected, evaluated and added to D. So the *posterior probability* $= P(f|D) = P(D|f) \times P(f)$ is used as an approximation of the objective function which means it is used as a surrogate objective function. A surrogate function should be relatively cheap to evaluate with a popular model being the Gaussian Process. The final piece is the acquisition function which is a technique used to select the next point x_{n+1} from the search space. The objective function is then evaluated at this point $f(x_{n+1})$ with both the point and result being added to D and used to updated the *posterior*. In summary:

1. Select a point by optimising the acquisition function.
2. Evaluate the point with the objective function.
3. Update D and the surrogate function.
4. Repeat.

Generally this process is repeated until some stop criteria are reached.

6.2.6 Feature Engineering

Feature engineering makes use of high level information in order to increase the performance of ML algorithms. General examples include taking the log of a variable or One-hot encoding which assigns each element e a unique value.

$$e \in [0, 3] \rightarrow 0 = (0, 0, 0, 1);\ 1 = (0, 0, 1, 0);\ 2 = (0, 1, 0, 0);\ 3 = (1, 0, 0, 0)$$

Particle physics examples include substructure variables and b-tagging. Feature engineering has the advantage of not treating DNNs as black boxes as domain knowledge is used to create the new features. However it is also hoped that the models can learn some new

high level features using this strategy. However, this typically requires huge data sets. Data preprocessing is important for DNNs to learn correctly. Networks require that each of the input features are normalised so as to have similar magnitudes and variance. This is to avoid unintentional biasing of certain features with higher values.

Chapter 7

Object Definitions

The objects used in the analysis chapter are shared by the various studies described in sections 8.1, 8.2, and 8.3. The basic principle is to use information from the detector to build low level objects, which can then be combined into higher level objects with some physical meaning such as leptons and jets. Depending on the regime of the analysis in question these objects can be directly combined, through the use of combinatorics, to re-create the analysis dependent physical objects that are being searched for or measured, to give context for studies mentioned later on these objects would be originating from massive particles e.g. top quarks (173 GeV), W bosons (80 GeV) and Higgs boson (125 GeV) decays. The p_T regime being considered is a further factor in building "high"-level objects as when re-constructing objects originating from resolved decays there tends to be larger ΔR separation between the constituents as such, combinatoric matching and multivariate algorithms are employed for reconstruction. However, there are cases when the objects decaying have a high p_T which will have highly collimated decay products with low ΔR separation, and in these cases boosted objects can be employed which effectively use larger jets. Because the two general purpose detectors at the LHC are ATLAS and CMS which are used to check and corroborate the physics searches and measurements between themselves physicists are not allowed to work on both. However, as they share the primary physics goals and have some similarities between their detectors there are some similarities between the two for example in how the objects are reconstructed e.g particle flow jets. The rest of this chapter will focus on how ATLAS reconstructs its objects and other algorithms it develops centrally for use by analyses. These algorithms are used to identify specific objects for example jets originating from b-hadron decays referred to as b-tagging. The ATLAS detector itself can't detect neutrinos (ν) however, the missing transverse momentum (P_T^{miss}) can be used to help identify particles that escape the detector such as neutrinos.

7.1 Low Level Objects

The low level objects described in this section are used to reconstruct both leptons and jets, where leptons typically refer to electrons (e) and muons (μ) and jets are the hadronic decay products typically from hadrons.

7.1.1 Calorimeter Topological Clusters

Energy deposits in the calorimeters are clustered together with ATLAS employing a topological clustering algorithm with them being referred to as Topo-clusters. The clusters are calibrated both on the electro-magnetic (EM) scale or on the continuous "local hadronic cell weighting" (LCW) scale. The former (EM) aims to reconstruct electrons and photons with the latter (LCW) aims to reconstruct signals with significant hadronic content such as jets and τ-leptons. The LCW calibrates clusters jointly using both the EM and hadronic scales weighting them with an EM likelihood for the given cluster.

Topo-clusters

Topological clusters are calorimeter cells grouped into clusters based on the neighbouring cells to the seed cell having a certain energy compared to the expected noise. It is split into two steps, the cluster maker and the cluster splitter. The cluster maker initially identifies all the possible seed cells with a threshold energy above four times the standard deviation of the expected noise referred to as cell signal significance.

$$Z_{cell}^{EM} = \frac{E_{cell}^{EM}}{\sigma_{noise,cell}^{EM}} > 4$$

Where, E_{cell}^{EM} and $\sigma_{noise,cell}^{EM}$ are measured on the EM scale. The expected noise is from both the electronics and pile-up which is dependent on the run conditions. Then adjacent cells are iteratively added, in both the lateral and longitudinal direction, if their energy exceeds two times the standard deviation of the expected noise ($Z_{cell}^{EM} > 2$). Finally all cells neighbouring the currently selected cells are also added to ensure that the tails of the energy deposits are included. This is then followed by the cluster splitting step which is designed to split clusters in half containing two local maxima. The clustering algorithm employed here aims to separate energy deposits of a different nature i.e electromagnetic and hadronic rather than identify deposits originating from particles [69].

7.1.2 Tracks

As a charged particle such as an electron travels through the detector it creates "hits" in the various subsystems thus leaving a "track" of its trajectory. Due to the magnetic

field present in the detector charged particles will have a curved trajectory from which both the charge and momentum can be inferred. The two main subsystems in the ATLAS detectors used for measuring tracks are the Inner Detector, located around the centre of the interaction point, and the muon systems which are the furthest sub-systems from the interaction point.

Inner Detector (ID) Tracks

Tracks in the inner detector are reconstructed from three dimensional "space-points" and are the result of combinatorial matching algorithms with each "proto-track" being seeded by at least three space-points. Space-points are the reconstructed hits in the various layers of the ID. In the pixel layers a space-point is defined from a single ID-cluster. In the SCT layers a space-point is defined from ID-clusters located on both sides of the strip in order to obtain a three dimensional measurement. The previously mentioned ID-clusters are formed from the energy deposits left by passing charged particles hitting a ID layer and are grouped with connected component analysis [70] where the energy deposits surpass a certain threshold energy.

The proto-tracks are then iteratively combined using a Kalman filter [71]. Clusters used by multiple tracks - are either identified as merged or shared but not both. With proto-tracks being ranked in terms of purity: SCT space-points only, pixel space-points only and mixed using space points from both. The final track candidates must then satisfy the following criteria:

- $p_T > 400$ MeV,
- $|\eta| < 2.5$,
- ≥ 7 pixel and SCT space-points,
- ≤ 1 shared pixel space-point or 2 shared SCT pixel space-point,
- ≤ 2 holes where the track passes through a layer (pixel or SCT) but there is no matching space-point,
- ≤ 1 hole in the pixel detector,
- $|d_0^{BL}| < 2.0$ mm,
- $|Z_0^{BL} \sin(\theta)| < 3.0$ mm.

Where d_0^{BL} is the transverse impact parameter calculated with respect to the measured beam-line position, Z_0^{BL} is the longitudinal difference along the beam line between between d_0^{BL} and the primary vertex and θ is the polar angle of the track.

Ambiguities arising from multiple track candidates sharing space-points are resolved by assigning each candidate a score and ranking them. Track scores are defined with: the number of matched space-points, the number of holes, a χ^2 track fit and the $\log(p_T)$ of the track. Holes negatively effect the track score while space-point multiplicity positively effect it, priority is given to high p_T tracks both to suppress incorrect space-point assignment prevalent at low p_T and because they are more likely to be interesting candidates for the physics analyses being conducted. In cases where space-points are matched to multiple tracks limits are placed on the number of shared space-points, if it's a pixel space-point a NN is used to separate and identify merged space-points from shared space-points. Due to the pixel being closest to the interaction point it is common for multiple charged particles to hit the same/neighbouring pixels creating merged ID-clusters which can be matched to multiple track candidates without penalty as is the case with shared space-points. For more details see [72].

Primary and Secondary Vertices

Vertices form an important part of reconstructing the underlying physics and are defined with the ID tracks. The primary vertex refers to the reconstructed hard proton-proton scattering event and is defined as the vertex candidate with the highest $\sum p_T^2$. Vertex candidates are identified through an iterative procedure and are seeded from the beam spot in the transverse plane. Combinations of tracks are selected to find the best possible vertex position identified through χ^2 minimisation. The process is repeated until all possible vertices are found and all tracks are associated with a vertex where possible. For more details about the primary vertex identification see [73].

Secondary vertices are aimed at jet flavour identification and are a key input to the b-tagging algorithms. Because of this only tracks contained within a specific jet cone are considered as the secondary vertex finder (SVF) algorithm finds all two track vertices, then tries to merge these two track vertices that are close together to create multi-track vertices. For full details about the SVF algorithm see [74].

Muon Spectrometer (MS) Tracks

In order to identify muons information is combined from several different subsystems namely: ID, MS and calorimeters. Tracks form a vital role in muon reconstruction with ID tracks (described previously) being combined with MS tracks. Tracks in the MS are built from hits in the MDT, RPC and TGC subsystems. Hits in the MDT are combined into segments using a straight line fit. Hits are identified using a Hough transform [75] which searches each MDT and adjacent trigger chamber for hits aligned on the bending plane of the detector. Hits in the RPC and TGC are used to measure the coordinate orthogonal to the bending plane. Segments in the CSC detectors are built with a different combinatorial

search algorithm in the $\eta - \phi$ plane and loosely requires tracks to be compatible with the luminous region.

The track candidates are then built from the hits segments in the different MS subsystems. They are seeded with segments from the middle layers and then combined with segments from the outer and inner layers. The segments are ranked based on hit multiplicity and fit quality with track candidates requiring at least two matched segments to build a track. An overlap removal algorithm is used to resolve tracks sharing segments which allows for tracks matched to three segments to share two of those segments as long as the outer layer segments for both share no hits. This is to ensure that close-by muons are still identified. The hits associated with each track candidate are fitted using a global χ^2 fit. For more details see [76].

7.2 Leptons

In ATLAS the two primary leptons reconstructed are electrons (e) and muons (μ). Although taus (τ) can be reconstructed it is not so straightforward to do so and in general require dedicated analyses. Neutrinos (ν) can't be directly detected by the ATLAS detector but the missing transverse momentum (P_T^{miss}) from the event can be used as an estimate and is described in Section 7.7. In the analyses described in Chapter 8 all tracks used in lepton identification must match the primary vertex of the event, $|Z_0^{BL}\sin(\theta)| < 0.5$ mm.

7.2.1 Electrons

Electrons are reconstructed by matching ID tracks with the topological clusters in the calorimeters, the path of an electron through the detector can be seen in Figure 7.1 illustrating this. If several track candidates can be matched to one cluster then the candidate is selected such that: its extrapolated trajectory is closest in $\eta - \phi$ space to the matching cluster barycentres, as measured in the 2nd layer of the calorimeter; the number of space-points in silicon layers, prioritising space-points in the innermost layer. A successful electron candidate will have a track with at least four space-points in silicon layers, a matching topological cluster in the calorimeter and no association with a vertex from a photon conversion.

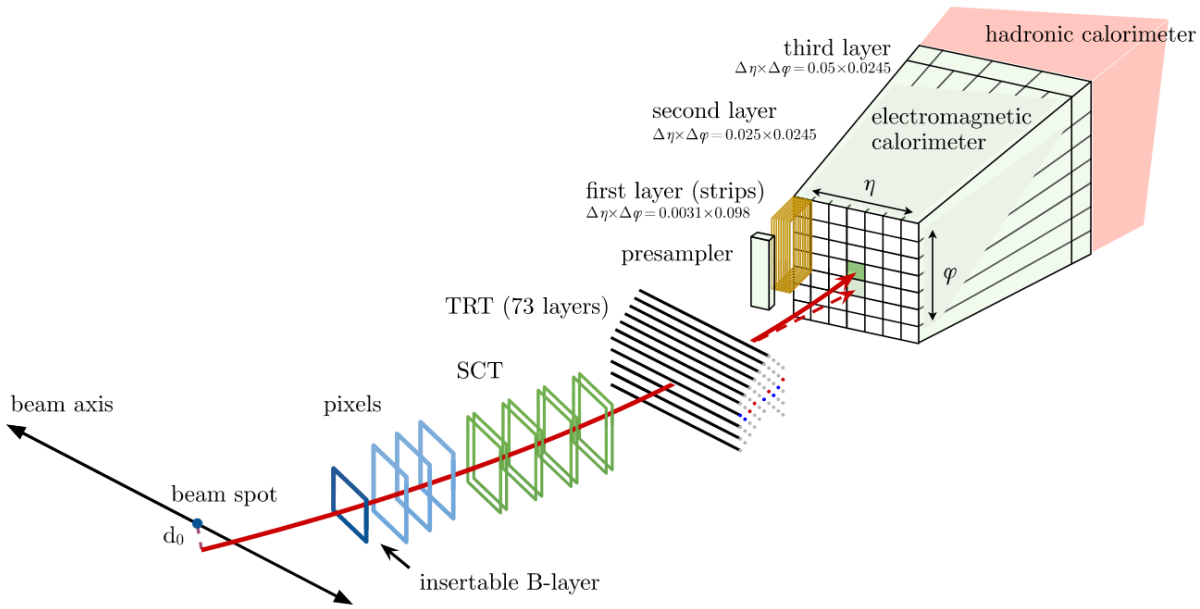


Figure 7.1: An illustration of an electron's trajectory (solid red line) through the detector's various sub-systems. The dashed red line shows a photon produced from the resultant EM shower [77].

To separate electrons from photons the topological clusters are initially tagged before track matching to either be a seed cluster for an electron or photon candidate with electron candidates requiring a cluster energy of ≥ 1 GeV and match to a satisfactory track. These candidate clusters are then combined with satellite clusters if they are within proximity in order to capture the full EM shower produced by bremsstrahlung radiation with these being referred to as superclusters.

To further aid the purity with which electrons are reconstructed the electrons are further assigned a tag indicating the confidence in their identification. It is based on a likelihood discriminant with information from the primary electron track, the lateral and longitudinal shape of the EM shower in the calorimeter and the spacial matching of the track to the cluster. It has three operating points Loose, Medium and Tight defined in terms of efficiency with targets of 93%, 88% and 80%. For more details see [77, 78].

Electrons used in the analyses described in chapter 8 are required to have $p_T > 10$ GeV, $|\eta| < 2.47$ with candidates from the barrel-endcap transition region ($1.37 < \eta < 1.52$) being excluded. They are also required to pass the Medium likelihood identification criteria.

7.2.2 Muons

Muons are reconstructed using information from ID, MS and the calorimeter. There are broadly 5 different strategies dubbed: combined (CB), inside-out combined (IO), muon-spectrometer extrapolated (ME), segment-tagged (ST) and calorimeter-tagged (CT).

For the CB muons MS tracks are matched to the ID tracks and fitted together based on the combined ID and MS hits while allowing for energy loss in the calorimeters. The trajectory from the combined track is used to further refine the hits in the MS. For $|\eta| > 2.5$ a subset of CB muons exists where MS tracks are matched to short track segments reconstructed purely from hits in silicon layers (pixel / SCT) called Silicon-associated forward muons (SiF).

IO muons are reconstructed using a complementary inside-out algorithm with the aim of finding muons where the MS has limited coverage or for muons with low p_T that don't reach the middle layers. It therefore doesn't use the full MS tracks described previously, instead it extrapolates the ID tracks to the MS and searches for ≥ 3 loosely aligned MS hits. These extrapolated ID MS tracks are then fitted together with the energy loss in the calorimeter.

ME muons are in a sense the reverse strategy, MS tracks that can't be matched to an ID track are extrapolated back to the beam line defining them. The aim is to take advantage of the full MS coverage of $|\eta| = 2.7$ which is outside of the ID coverage.

ST muons are defined by extrapolating ID tracks to the MS and requiring that they satisfy tight angular matching to ≥ 1 MS segment. If successfully matched it is used to define a muon with its parameters being defined directly from the ID track fit.

CT muons are identified by extrapolating ID tracks through the calorimeters and searching for energy deposits consistent with a minimum-ionising particle. It is this information that is used to tag the ID track as a muon with the muon parameters taken directly from the ID track fit. There is a p_T criteria of 5 GeV for the ID tracks used here to account for background contamination at low p_T as opposed to 2 GeV for the previous methods.

On top of these several Working Points (WPs) are defined based on information from their reconstruction, namely: hit multiplicity in both the ID and MS subsystems, the track fit properties and the variables used to assess the individual hits in either the ID or MS. There is a distinction to be made between prompt and non-prompt muons, where prompt muons originate from the primary vertex and non-prompt muons are the result of some hadron decaying in-flight. Distinguishing non-prompt muons originating from heavy hadron flavour decays (b, c) from those of lighter flavours is important as these can be well separated from prompt muons.

The standard WPs for the muon identification criteria are Loose, Medium and Tight and are designed with different purposes. The term "precision station" refers to the number of MS stations where the candidate has at least 3 hits in the MDT or CSC sub-systems.

The Loose WP accepts CB and IO muon definitions as well as ST and CT muon definitions in the $|\eta| < 0.1$ range to account for the gap in the MS. To extend outside the ID coverage ME and SiF muons are also included in the range $2.5 < |\eta| < 2.7$ and

are required to have at least three precision stations. It was optimised for $H \to \mu\mu\mu\mu$ reconstruction which due to the multiplicity benefits from higher efficiency at the cost of larger systematics and less purity.

The Medium WP accepts primarily CB and IO muons which are required to have at least two precision stations apart from candidates in the $|\eta| < 0.1$ range where only one is required. It also accepts ME and SiF muons in similar fashion to the Loose requirement in order to extend the coverage. It is suitable for most physics analyses and was optimised to keep the systematic uncertainties in the prompt-muon efficiency and background rejection small.

The Tight WP uses only CB and IO muons with at least 2 precision stations. It then further requires that their combined tracks normalised χ^2 fit is less than 8 in order to reject tracks due to in-flight hadron decays. It was optimised for analyses that are limited by backgrounds from non-prompt muons therefore it has an even higher purity with only a slight efficiency lost compare to the Medium WP. For more information see [79].

The muons used in the analyses described in chapter 8 are required to have $p_T > 10$ GeV and $|\eta| < 2.5$ and only use the CB, IO, ST and CT muon definitions. They are also required to satisfy the Loose identification criteria and have $|d_0^{BL}| < 3.0$.

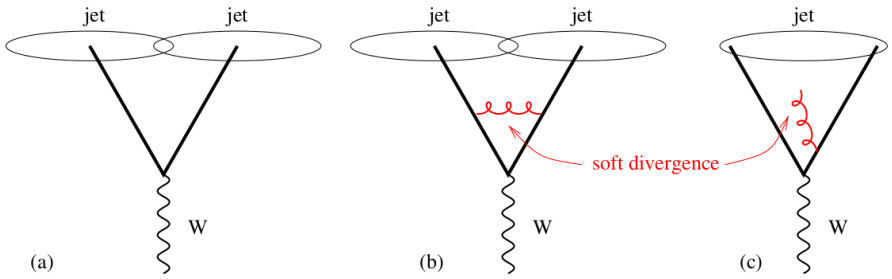
7.3 Jets

When quarks and gluons are produced at the LHC they will almost immediately hadronise due to colour confinement, that is with the exception of the top quark which due to its large mass decays even faster than it can hadronise. Because of this and the subsequent interactions with the detector, the initial hadrons end up producing a collimated spray of hadrons that are reconstructed as jets. Due to the messy signatures left by hadrons especially when compared to electrons, muons and photons there are many strategies that can be used to reconstruct them. Because of this the "Snowmass accord" [80] was set which requires that jet definitions adhere to the following requirements:

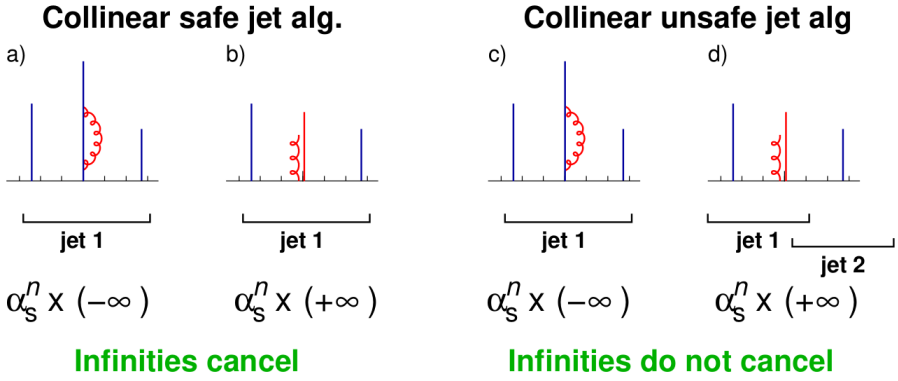
- Simple to implement in an experimental analysis,
- Simple to implement in the theoretical calculation,
- Defined at any order of perturbation theory,
- Yields finite cross-sections at any order of perturbation theory,
- Yields a cross section that is relatively insensitive to hadronisation.

An important consequence of the last point is that jet algorithms must be both infra red and collinear safe, which means that if an event is modified by either an additional

soft emission or a collinear splitting, the overall set of hard jets found in the event remains unchanged.



(a) Example of an infra-red emission and modifiocation of jet multiplicity. [81]



(b) Example of a collinear splitting and modifiocation of jet multiplicity. [81]

Figure 7.2

Cone type jet algorithms notoriously suffer from one or both of these issues. From Figure 7.2 we can see how a soft divergence 7.2a or a collinear emission 7.2b can modify the overall number of jets found in the event. Sequential recombination algorithms behave much better but can produce irregular jet shapes. The three most commonly used algorithms can be generalised as such:

- $d_{ij} = min(p_{ti}^{2p}, p_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}$,
- $d_{iB} = p_{ti}^{2p}$,
- $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$.

The parameter p can take one of three values 1, 0, -1 with the algorithms being known as the: k_t ($p = 1$) and clusters from the softest candidates first; Cambridge Aachen ($p = 0$) where the energy dependence has been removed [82]; and anti-k_t ($p = -1$) which clusters from the hardest candidates first [83]. Their basic steps are the same:

1. Find all d_{ij} and d_{iB}.

2. Iterating through each candidate check which satisfies $min(d_{ij}, d_{iB})$.

3. If it's d_{ij} combine i and j into one and return to step 1.

4. Else if it's d_{iB}, than declare i to be a final state jet and remove it from the set of candidates.

5. Stop when no candidates remain.

The anti-k_t algorithm has the useful property of producing regular shaped jets as well as being IRC safe. Because of this it is the default jet clustering algorithm used in the analysis chapter 8. However, due to it clustering from the hardest candidates first its clustering history is not particularly useful for deriving substructure variables. The shapes of the various algorithms can be seen in Figure 7.3.

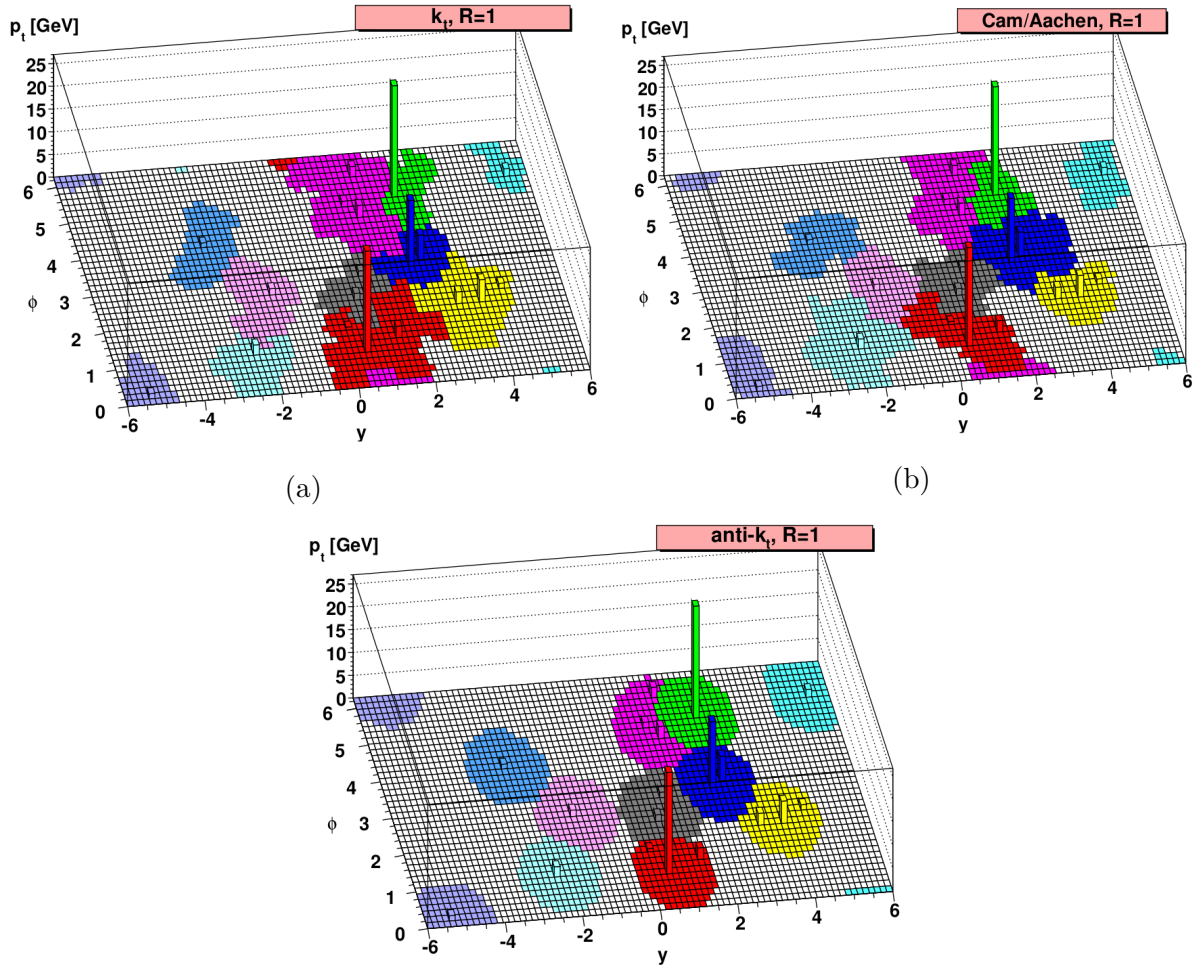


Figure 7.3: A parton level event and soft emissions generated with Herwig, clustered with the three sequential clustering algorithms to illustrate their shapes. With k_t in Figure 7.3a, Cambridge / Aachen in Figure 7.3b and anti-k_t in Figure 7.3c [81].

7.3.1 EMtopo jets

EMtopo jets are the primary type of jet being referred to throughout this text as either jets or small-jets and are used by the studies in sections 8.1 and 8.2, they are required to have $p_T > 25$ GeV and $|eta| < 2.5$. EMtopo jets are reconstructed solely using calorimeter information. They are defined with the *anti-*k_t algorithm with a radius parameter $R = 0.4$ implemented with the FastJet 3.2.2 software package [84, 85]. It uses calorimeter energy deposits, whose cells are first clustered into topo-clusters described in Subsection 7.1.1, as inputs. These clusters are calibrated at the EM-scale and are origin corrected.

7.3.2 Particle Flow (PFlow) jets

PFlow-jets use an alternative algorithm to the EMtopo-jets previously described, it is the new recommended jet collection by ATLAS and is used in the Legacy analysis studies in Section 8.3, as with EMtopo-jets they are clustered with the *anti-*k_t algorithm ($R = 0.4$) and are required to have $p_T > 25$ GeV and $|\eta| < 2.5$. It combines information from the momentum measurements in the ID (ID tracks) and energy deposits in the calorimeter (topo-clusters). The main advantages of the PFlow algorithm is that it reduces the effects of "pile-up" particularly at lower p_T thus improving energy and angular resolution as well as its reconstruction efficiency compared to EMtopo-jets. It also maintains the performance at high p_T achieved by the EMtopo algorithm due to the excellent calorimeter performance at high energies.

Pile-up is the term used to describe the resulting signals from other interactions typically not of interest resulting from either soft-scattering interactions from the same bunch crossing or residual signal remnants from other bunch crossings.

The PFlow algorithm uses a subtraction algorithm to remove overlaps for an object between its momentum and energy measurements made by the ID and calorimeter respectively, this avoids double counting a particles energy. Initially a set of well-measured tracks are created from the inner-detector. These are then matched to topological-clusters in the calorimeter. Then once the optimal configuration of matchings for tracks to clusters is found (one track can be matched to many clusters) the energy of the track is removed cell by cell from the set of matched clusters. Finally if an objects (matched track to clusters) energy remnants are consistent with the expected shower fluctuations of a single particle then it is assumed that these energy deposits came from a single particle, are a result of showering in the calorimeter, and are also removed. After this the set of selected tracks and remaining topological-clusters should ideally represent the reconstructed event. A flow chart of the PFlow algorithm and how ID-tracks and calorimeter clusters are combined can be seen in Figure 7.4, for more details see [86].

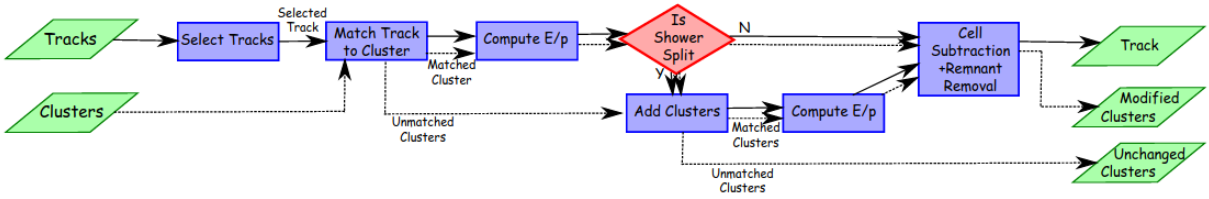


Figure 7.4: Flow chart of the PFlow algorithm showing how tracks and clusters are combined [86].

It should be noted that the PFlow algorithm has been updated since [86] using the updated simulation for Run 2. Both the expected mean value of the energy deposited by pions and the shower profiles were recalculated. In addition an update to the transition algorithm between track and calorimeter energies targeting high p_T jets. Previously a simple $p_T < 40$ GeV cut on a track was used where in those cases it was not subtracted from the calorimeter cluster. This is due to these cases predominantly falling in the centre of the jet which has the consequence that they are poorly isolated from the surrounding clusters which make it difficult to accurately subtract their energies from that of the calorimeters. However, now a more sophisticated approach is used with tracks selected if $p_T < 100$ GeV then if the energy E^{clus} in a cone ($\Delta R = 0.15$) around the extrapolated particle satisfies:

$$\frac{E^{clues} - \langle E_{dep} \rangle}{\sigma(E_{dep})} > 33.2 \times log_{10}(40 \text{ GeV}/p_T^{trk})$$

then the subtraction is not performed. This takes into account the energy density of the clusters and effectively slowly truncates the algorithm but still allows for some subtraction at high energies where the calorimeter performs well, for further details see [87].

7.3.3 Jet Energy Calibration and Resolution

The resultant jet collections are then calibrated as well as having their energy resolutions measured. Both of these quantities and their associated systematic uncertainties are important to understand for physics analyses and are derived by matching the reconstructed jets to particle level jets, derived from the MC truth information, with a $\Delta R < 0.3$ requirement and using $36 - 81 fb^{-1}$ of data collected between 2015-2017.

Jet energy scale (JES) calibration

The purpose of the jet energy scale calibration is to restore the jet collections energy so that they match particle level jets derived from the MC truth information. The sequence of corrections is shown in Figure 7.5 and is the same for EMtopo and PFlow jets. For JES calibrations dijet MC events are used.

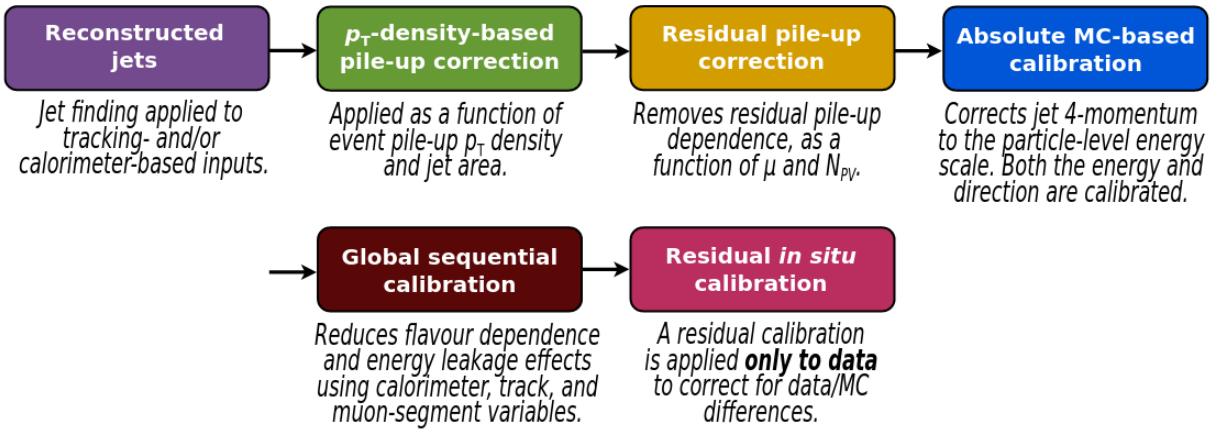


Figure 7.5: Flow chart of the sequence of correction for the jet energy scale (JES) calibrations [87].

The initial p_T based pile-up correction removes the excess energy in the event from additional pp collisions from the same bunch crossing (in-time) or from nearby (out-of-time) bunch crossings. The pile-up corrections can be summarised in two steps corresponding to the green and yellow bubbles seen in Figure 7.5.

The p_T density based pile-up correction initially uses the area A of a jet to assess the degree to which it is affected by pile-up. It is determined by counting how many ghost tracks [88] are associated with a jet after clustering. After, the pile-up contribution is estimated using the median p_T density $\rho = p_T/A$. For this jets are clustered with the k_t algorithm ($R = 0.4$) in the range $|\eta| < 2$ which limits the computation to the central region of the detector in order to better measure pile-up activity due to $\rho \to 0$ for $|\eta| > 2$. The k_t algorithm is used because it naturally reconstructs jets as well as the soft background. The variable ρ is used because it reduces the bias from high p_T events that populate the tails.

The residual pile-up correction is aimed at describing the pile-up effects present in the forward calorimeter region and in the busy cores of high p_T jets where ρ doesn't perform well. Because of this, without the residual corrections, there is an observable dependence of the anti-k_t jet p_T on pile-up activity. To address this the residual dependence is defined to be the difference between the reconstructed jet's p_T and the particle level jets p_T derived from the MC truth information and is defined below:

$$P_T^{corr} = p_T^{reco} - \rho \times A - \alpha \times (N_{PV} - 1) - \beta \times \mu,$$

where p_T^{reco} is defined before any pile-up corrections, N_{PV} is the number of primary vertices in the event and μ is the mean number of interactions per bunch crossing. Here α and β are coefficients derived from a separate linear fits binned in-terms of p_T^{true} and $|\eta_{det}|$.

The absolute JES and η calibrations correct the jets four-momentum to the particle-

level energy scale which accounts for non-compensating calorimeter response, energy losses in passive material, out-of-cone effects and biases in the reconstructed jets η. These biases are typically a result of the different calorimeter technologies being stitched together. For this calibration anti-k_T ($R = 0.4$) jets are from a Pythia 8 dijet MC sample post pile-up corrections. Once the particle-level jets are matched to the reconstructed jets the average jet energy response R is defined by fitting a Gaussian to the E^{reco}/E^{true} distribution measured in bins of E^{true} and η_{det}. The jet calibration factor is derived from the inverse of the average energy response.

A further η bias can be observed in the reconstructed jets when compared to the particle-level jets derived from the MC truth information. It is most observable in jets that span two different calorimeter regions with different energy responses due to either a different geometry or technology. The effect of this is that the reconstructed jet will have an energy imbalance internally thus altering its four-momentum. Jets reconstructed around the barrel-endcap transition region ($|\eta| \approx 1.4$) and the endcap-forward region ($|\eta| \approx 3.1$) are particularly effected. So a correction is derived between the reconstructed jet's η and the particle-level jet's η. This correction only alters the the reconstructed jet's p_T and η and does not effect the full four-momentum.

Global sequential calibration aims to reduce fluctuations in reconstructed jets caused by either being initiated by a quark or gluon and improve the jet energy resolution whilst maintaining the average jet energy response. Jets initiated by quarks have a tendency to consist of hadrons that possess a higher fraction of the jet p_T and so penetrate further into the calorimeter. On the other hand jets initiated by gluons tend to consist of more hadrons with a lower p_T fraction which has the effect that these jets have a lower calorimeter response and a wider transverse profile. The corrections are based on jet observables: the longitudinal structure of the energy deposits in the calorimeter, associated tracking information and information from the MS directly behind the jet. The corrections to the four-momentum, derived as a function of p_T^{true} and $|\eta_{det}|$, improve six observables that improve the JES with each observable being corrected independently of each other. The six observables are:

- (PFlow only) $f_{charged}$ the fraction of the jet p_T measured from ghost-associated tracks with $p_T > 500$ MeV ($|\eta_{det}| < 2.5$);
- f_{tile0} the fraction of jet energy measured in the first layer of the hadronic Tile calorimeter ($|\eta|) < 1.7$;
- f_{LAr3} the fraction of jet energy measured in the third layer of the electromagnetic LAr calorimeter ($|\eta_{det}| < 3.5$);
- n_{trk} the number of tracks with $p_T > 1$ GeV ghost-associated with the jet ($|\eta_{det}| < 2.5$);

- w_{trk} a measure of the track width done by taking the transverse distance in the $\eta - \phi$ plane, weighting it by the mean p_T, between the jet axis and all tracks of $p_T > 1$ GeV ghost-associated with the jet ($|\eta_{det}| < 2.5$);

- $n_{segments}$, the number of muon track segments ghost-associated with the jet ($|\eta_{det}| < 2.7$).

The residual in situ calibration is the final step and aims to account for differences between jet response in data and simulation. These are primarily due to imperfect simulations of the initial hard scattering process along with the rest of the event, jet formation, pile-up and the particle interactions within the detector. Corrections are applied in three stages with the initial η calibration correcting the energy scale of jets in the forward region ($0.8 \leq |\eta| < 4.5$) to match jets in the central region ($|\eta| < 0.8$). This is done by balancing the p_T of a jet against a well defined object using dijet events. Secondly, the hadronic recoil in an event is balanced against that of a well calibrated Z boson or photon, which are the standard candles of particle physics, using Z+jet and γ+jet events respectively. Finally, high p_T jets are calibrated with multijet events using well calibrated low p_T jets. For more details see [87].

Jet energy resolution (JER) measurement

When analysing physics observables that decay to jets such as Higgs decays, top quarks and W/Z bosons having an accurate understanding of the jet energy resolution is important. The JER also has an affect on the missing transverse momentum in an event and can affect physics searches / measurements using neutrinos. It depends on the noise from the electronics and pile-up; the stochastic fluctuations in the amount of energy being deposited; and fluctuations in the amount of energy being deposited in passive material, which is not a function of p_T as opposed to the first two. To measure the JER the momentum must first be precisely measured which is done using a well understood dijet system where jets can be balanced against each other in similar fashion to how JES corrections are derived. The noise due to electronics and pile-up is measured from data using random unbiased triggers. For more details see [87].

7.4 *b*-tagging

Identifying *b*-hadron decays within ATLAS is important for many analyses, because of this algorithms are developed centrally by ATLAS. The two algorithms used in chapter 8 are MV2c10 used in Sections 8.1 and 8.2, paired with EMtopo jets described in Subsection 7.3.1, and DL1r used in Section 8.3, paired with PFlow jets described in Subsection 7.3.2. Both algorithms are machine learning based with MV2c10 using a boosted decision tree

(BDT) while the newer DL1r uses a feed forward neural network (DNN) that makes use of a recurrent neural network (RNNIP). However, the input features and fundamental identification criteria being used to identify b-hadrons are broadly similar. The algorithms work on 3 key aspects unique to b-hadrons. The first is the relatively long life time of b-hadrons which can cause a distinctive separation between the secondary vertex (from the b-hadrons decay) and the primary vertex (from the hard scattering event). The second is that b-hadron decays can have 2 vertexes themselves (excluding the primary vertex from the hard scattering event), where the inclusive vertex would be within the reconstructed jet itself. The third bit of information used is the topological structure of weak b-hadron and c-hadron decays and tries to reconstruct the full b-hadron decay chain.

The algorithms that exploit these features of b-tagging are outlined in the following from which these high level b-tagging algorithms (MV2c10 / DL1r) are trained. For the studies in Section 8.1 the algorithms were used in the 2016 configuration [89]. For the studies outlined in sections 8.2 and 8.3 the algorithms were updated with the availability of more data (2015-2017) [90, 91].

Impact Parameter Algorithms (IP2D) and (IP3D)

The IP2D and IP3D algorithms aim to exploit the long lifetime of b-hadrons to identify them. To do this they make use of the IBL added between run 1 and 2 specifically to improve b-tagging performance. The IP2D tagger makes use of the transverse impact parameter significance, d_0/σ_{d_0}, while the IP3D also utilises the longitudinal impact parameter significance, $z_0 \sin(\theta)/\sigma_{z_0 \sin(\theta)}$, along with the transverse impact parameter significance.

Secondary Vertex Finding Algorithm (SV1)

The secondary vertex algorithm aims to reconstruct the secondary displaced vertex within a jet leveraging the fact that b-hadrons can have 2 decay vertices. This is done through a χ^2 fit of all potential two-track vertices compatible with the jet. For details see [74].

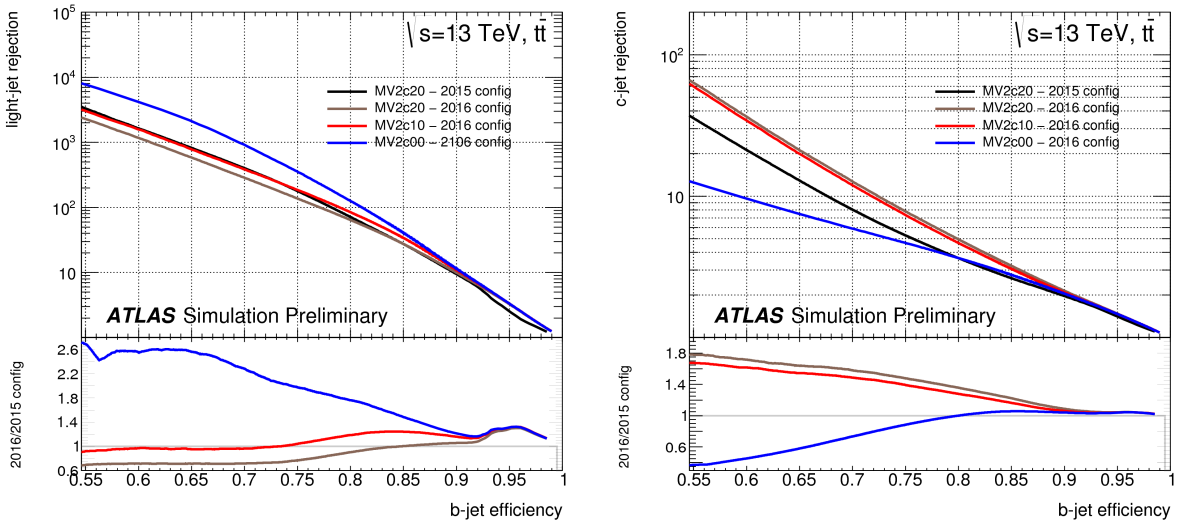
Decay Chain Multi-Vertex Algorithm (JetFitter)

The JetFitter algorithm exploits the topological structure of weak b-hadron and c-hadron decays inside the jet and uses a modified Kalman filter [71] to find a line which passes through the primary, b and c decay vertices thus approximating the b-hadron flight path. For details see [92].

7.4.1 MV2c10

The MV2c10 algorithm [89] uses the previously outlined lower-level algorithms (IP2D, IP3D, SV1 and JetFitter) utilising the unique properties of b-hadron decays. The outputs

of these algorithms are then combined in a BDT along with the kinematic jet variables p_T and η for the best performance. MV2c10 stands for multivariate algorithm, second iteration, where the c10 indicates the c-hadron background fraction used in training (here 7% and rounded up). This is important as the performance of the c- verses light-flavour jet rejection rate directly depends upon the background sample composition. It was trained on a sample of jets from $t\bar{t}$ events with b-jets set as the signal and the background consisting of c- and light-flavoured jets (with a composition of 7% and 93% respectively). There has been an improvement over the 2015 configuration MV2c20 algorithm in the c-jet rejection rate of approximately +40% with the light-jet rejection rate staying approximately constant (see Figure 7.6).



(a) (b)

Figure 7.6: Comparison of light-flavour jet rejection (7.6a) and c-jet rejection (7.6b) of the 2015 and 2016 MV2 configurations, taken from [89].

Working points are defined in terms of b-jet efficiency and have been set so that the loose has an efficiency of 85%, medium (77%), tight (70%) and very tight (60%). These are shown in Table 7.2 along with the BDT output and c-/light-jet rejection rates. The overall performance of the MV2c10 algorithm can be seen in Figures 7.7 and 7.8. In Figure 7.7 the jet composition per BDT output is shown. In Figure 7.8 the c-/light-jet rejection as a function of p_T is shown. At high BDT output scores light-flavour jets (consisting of up down and strange hadrons) are strongly suppressed (well separated), however, this is not the case for c-jets where the separation power is not as good meaning that c-jets are more likely to be mistagged as b-jets.

BDT Cut Value	b-jet Efficiency (%)	c-jet Rejection	Light-jet Rejection
0.9349	60	34	1538
0.8244	70	12	381
0.6459	77	6	134
0.1758	85	3.1	33

Table 7.1: Working points for the MV2c10 b-tagging algorithm along with the BDT cut values and c-/light-jet rejection taken from [89].

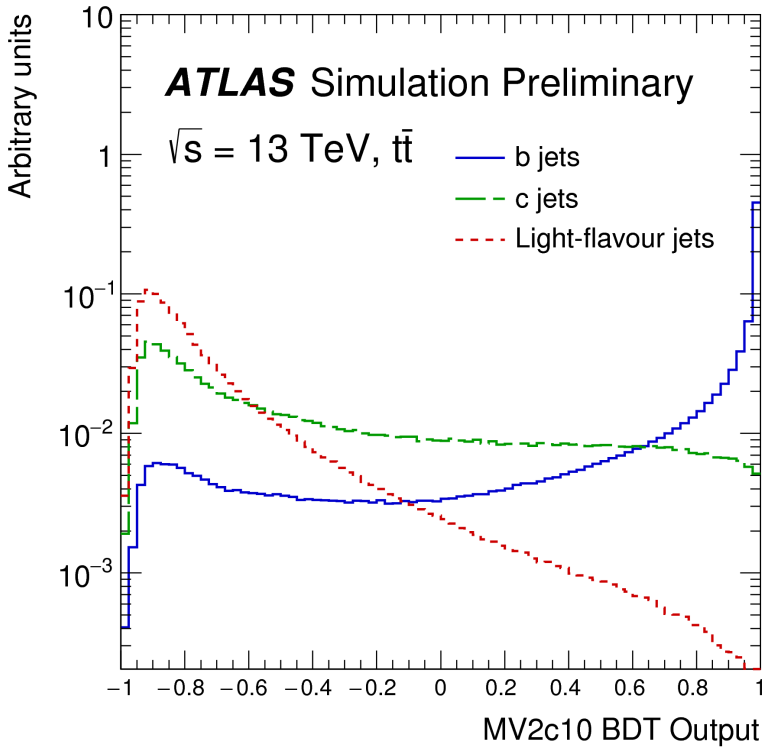
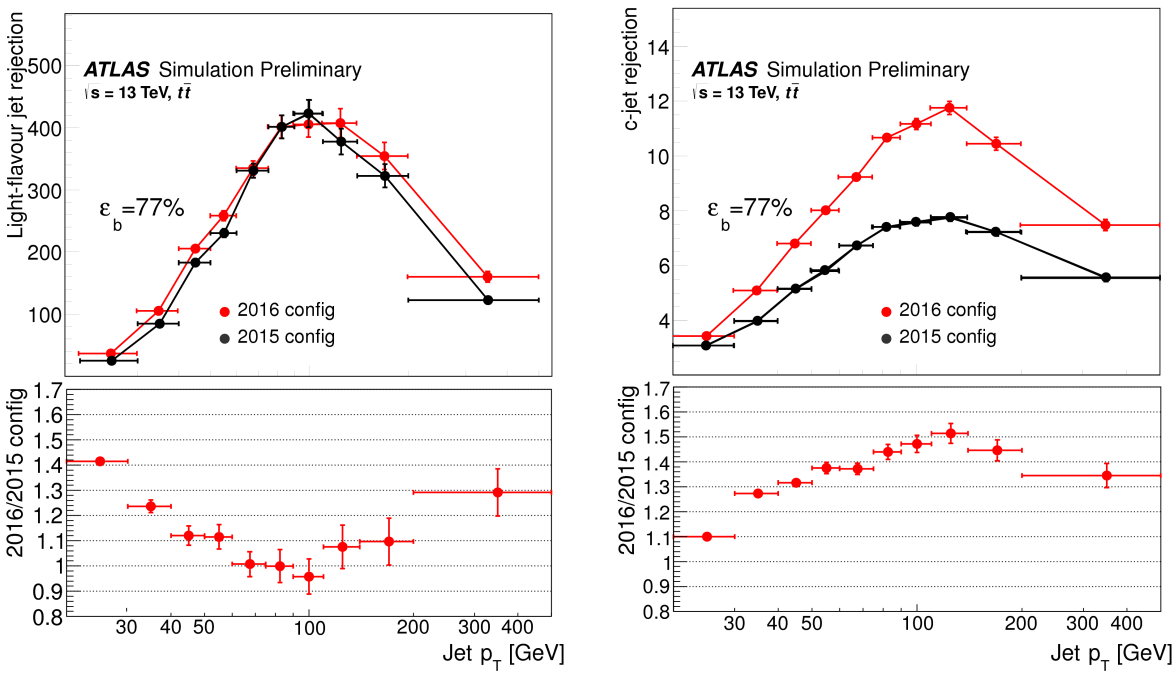


Figure 7.7: MV2c10 BDT output for b-, c- and light-flavoured jets evaluated with $t\bar{t}$ events, taken from [89].



(a) (b)

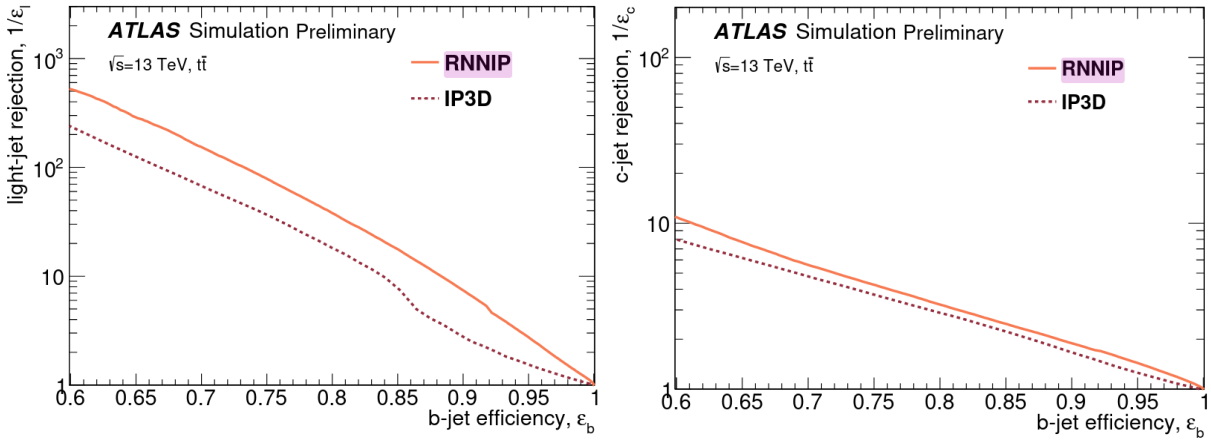
Figure 7.8: Light-flavour (7.8a) and c-jet (7.8b) rejection as a function of jet p_T, with each bin of the p_T set to the b-tagging efficiency of 77%, taken from [89].

7.4.2 DL1r

The DL1r b-tagging algorithm [90, 91] is the updated version and the current ATLAS recommendation that makes use of the same input features as MV2c10 with a few key differences and is the combination of the DL1 (DNN) and the RNNIP (RNN). The DL1 DNN is a simple feed forward network using the same inputs as the MV2 algorithm: IP2D, IP3D, SV1 and JetFitter as well as the kinematic variables jet p_T and η. It also makes use of the JetFitter c-tagging algorithm, which helps with c-jet rejection and was not included in the training of MV2.

The RNNIP is a recurrent neural network (RNN) track-based tagger that combines the two impact parameter algorithms IP2D and IP3D which use likelihood templates to compute per-flavour conditional likelihoods. The IP3D algorithm is limited by the assumption that each track is independent from all other tracks which is necessitated by the large sample sizes required to calculate it. The use of RNNs can overcome this challenge because they can learn the sequential dependencies between the tracks in a given jet. The transverse and longitudinal impact parameters along with the track / jet p_T fraction, angular distance between the track and jet-axis and the track-grade are fed into the network. A performance comparison of the RNNIP and IP3D can be seen in Figure 7.9 in terms of b-jet efficiency using $t\bar{t}$ events which shows a clear increase in

performance for RNNIP for *c*-jet and *light*-jet rejection.



(a) (b)

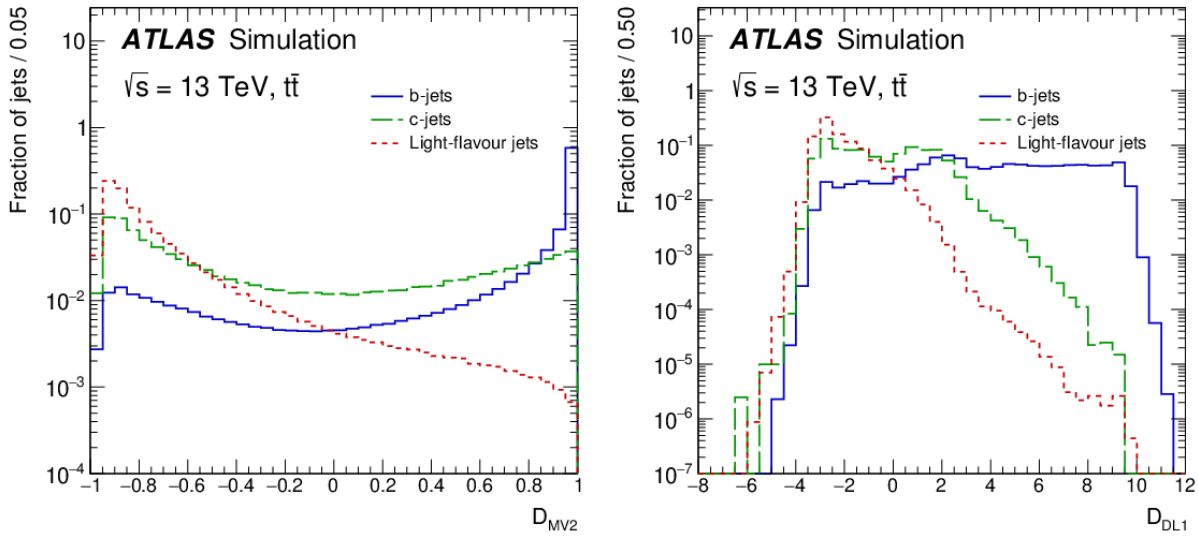
Figure 7.9: Comparison of RNNIP vs IP3D for *light*-flavour jet 7.9a and *c*-flavour jet 7.9b rejection as a function *b*-tagging efficiency for $t\bar{t}$ events [91].

DL1r uses all the previously mentioned DL1 inputs as well as the 3 output probabilities $Prob(b)$, $Prob(c)$ and $Prob(light)$ from the RNNIP for training, with both it and the RNNIP being trained separately on the same jet collection. This combination gives DL1r a slight performance gain over DL1.

Both DL1 and DL1r have similarly defined WP to the MV2 algorithms in terms of *b*-jet efficiency: loose has an efficiency of 85%, medium (77%), tight (70%) and very tight (60%). The performance comparisons between DL1 and MV2 can be seen below with: their WP for *c*-jet and *light*-jet rejection being shown in Table 7.2; their output discriminant being shown in Figure 7.10; and their *b*-tagging rejection of *c*-jets and *light*-jets distributions in Figure 7.11.

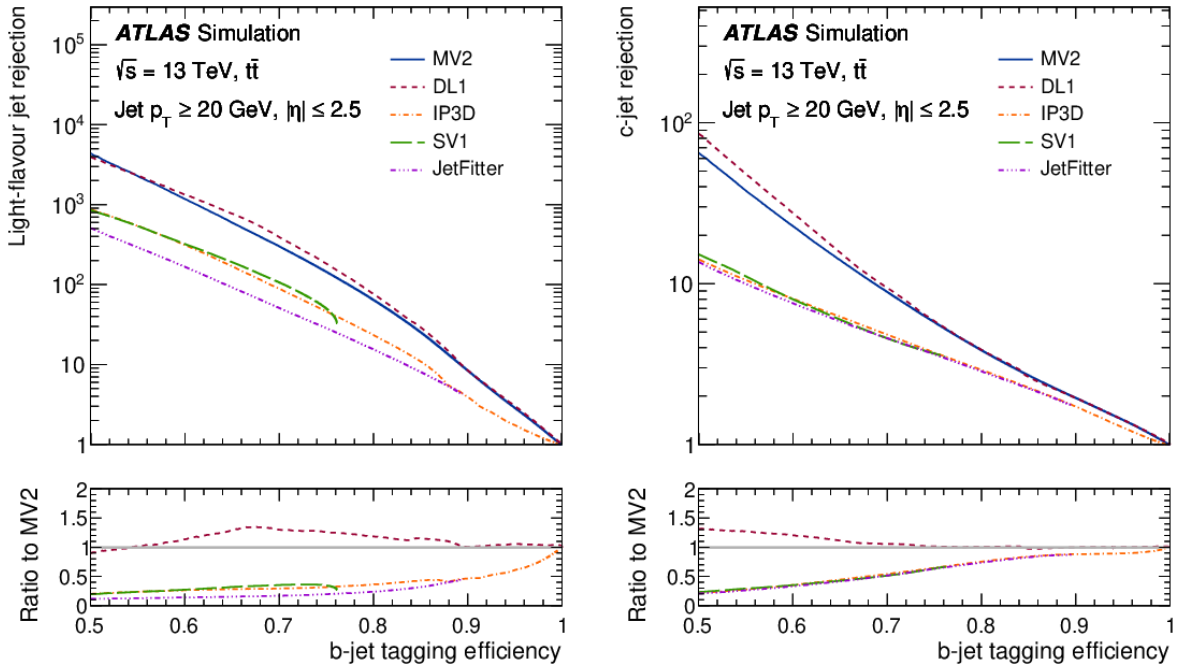
ϵ_b	MV2				DL1			
	Selection	Rejection			Selection	Rejection		
		c-jet	τ-jet	Light-flavour jet		*c*-jet	τ-jet	Light-flavour jet
60%	> 0.94	23	140	1200	> 2.74	27	220	1300
70%	> 0.83	8.9	36	300	> 2.02	9.4	43	390
77%	> 0.64	4.9	15	110	> 1.45	4.9	14	130
85%	> 0.11	2.7	6.1	25	> 0.46	2.6	3.9	29

Table 7.2: Output discriminant cuts used to define *b*-jet tagging efficiencies and their corresponding *c*-, *light* and τ-jet rejection for DL1 and MV2 [90].



(a) (b)

Figure 7.10: Distributions of the output discriminant of 7.10a MV2c10 and 7.10b DL1 *b*-tagging algorithms for *b*-, *c*- and *light*-flavour jets [90].



(a) (b)

Figure 7.11: The 7.11a light-flavour jet and 7.11b c-jet rejections versus the b-jet tagging efficiency for the IP3D, SV1, JetFitter, MV2 and DL1 *b*-tagging algorithms [90].

7.5 Boosted Jets

When dealing with high p_T objects it can be beneficial to model them with a single "big" jet due to the collimated nature of their decay products. To do this the same jet clustering algorithms are used (such as anti-k_t), as with standard jets previously described in Section 7.3, with $\Delta R = 1$ (rather than $\Delta R = 0.4$) so as to encapsulate their decay products. The primary benefit of this method is that various substructure type variables may be derived and assigned to these "big" jets which provide a certain discrimination power and aid in jet identification. There are two primary methods employed to model "big" jets they can either directly be clustered using the calorimeter information, similarly to EMtopo jets, referred to as Large Radius jets (LR-jets) or they can be reclustered using the standard ($\Delta R = 0.4$) jets as input constituents, these could either be the EMtopo jets in Subsection 7.3.1 or the PFlow jets in Subsection 7.3.2, and are referred to as reclustered jets (RC-jets). These jets are typically used to reconstruct boosted heavy objects such as the top quark, Higgs boson, W and Z boson with a boosted top quark candidate being shown in Figure 7.12. Because the decays of these objects are collimated to other relatively heavy objects (i.e not leptons and light quarks) there can be certain structures that show up in the energy deposits in the calorimeter which is what is referred to as substructure.

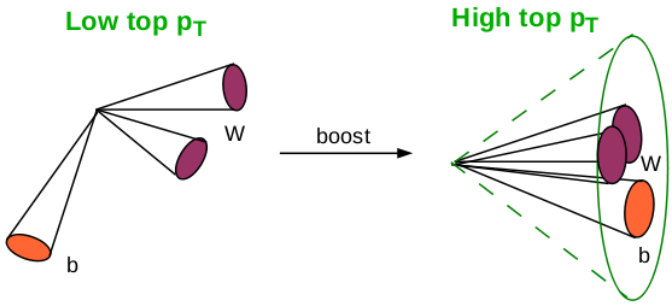


Figure 7.12: Resolved (left) vs collimated boosted top decays (right). The individual decay products of the resolved top decay are reconstructed as separate jets while the collimated boosted top decay is encompassed in one large jet.

7.5.1 Large Radius Jets (LR-jets)

Large radius jets are clustered directly from the calorimeter topo-clusters, in similar fashion to the standard $\Delta R = 0.4$ EMtopo jets, with the *anti-k_t* algorithm with $\Delta R = 1.0$ and are calibrated to the LCW-scale. LCW stands for the local cluster weighting calibration method which classifies the topological clusters on a continuous scale to be either electromagnetic or hadronic using their shower shapes and densities. An extra step that is required compared to the standard jets is a trimming procedure which removes soft radiation present from either the underlying event, pile-up or initial state radiation. This is done by taking the jet constituents (topo-clusters) and clustering them with the k_t al-

gorithm into constituent jets with $\Delta R = 0.2$ which will then be removed from the LR-jet if their momentum fraction is less than the threshold ($p_T/p_T^{LR} < 0.05$).

The JES calibration for LR-jets is done similarly to the standard $\Delta R = 0.4$ jets, using a PYTHIA MC sample, with a key difference being that no explicit off-set correction for pile-up contributions based on the number of primary vertices is done. LR-jets have a different set of systematics ascribed to them compared to the standard jets. They are calculated by comparing the LR calorimeter jets to LR track jets (whose systematics come from the ID) which overall show good spatial agreement between the two indicating no misalignment between the ID and calorimeters. Relative uncertainties are calculated from the ratio (r), for a given variable, between the track and calorimeter jets matched with $\Delta R < 0.3$. In the case of p_T (done for mass as well) the ratio would be:

$$r^{p_T} = \frac{p_T^{jet}}{p_T^{trackjet}},$$

then a double ratio (rr) is computed between data and MC in order to evaluate the agreement, for p_T:

$$rr^{p_T} = \frac{r_{data}^{p_T}}{r_{MC}^{p_T}}.$$

The mean values for the ratio r are expected to be well described, assuming that the detector is well modelled, because potential unaccounted for physics effects should affect the track and calorimeter jets similarly. The double ratio rr are used to test the data / MC agreement for jet p_T and m and allows for the calculation of an uncertainty associated with the calibration.

LR-jets are typically required to have a $p_T > 200$ GeV, $|\eta| < 2.0$ and $m > 50$ GeV although specific analyses may require tighter constraints. The tighter η constraints are to ensure that the whole jet is fully contained within the ID.

7.5.2 Reclustered Jets (RC-jets)

Reclustered jets are one of the methods used to model boosted objects. They use the standard $\Delta R = 0.4$ jets as inputs to the anti-k_t algorithm and reclusters them with $\Delta R = 1.0$ into a "big" jet. Doing this has several advantages, the primary being it allows direct propagation of the systematic uncertainties from the standard jets thus making combined measurements with resolved regions straight forward. It also leverages all the work that has gone into calibrating the standard jets and requires no extra studies. Because the jet algorithms rely on calorimeter information the detector simulation (see Section 5.5) used is important with full detector simulation being required for accurate substructure modelling. However, requiring full detector simulation is rather CPU costly and is a limiting factor for many analyses. Another advantage of RC-jets is that it provides a

way to have some substructure type variables available while using the AF2 fast detector simulation which utilises FastCaloSim for the calorimeter simulation. Because of this the true substructure variables of both standard $\Delta R = 0.4$ jets and $\Delta R = 1.0$ LR-jets that rely on precise calorimeter clustering are poorly modelled. Using RC-jets circumvents this to a degree as the the standard $\Delta R = 0.4$ jets four-vector is calibrated for AF2 meaning that any variables that can be defined with these inputs and so do not directly rely on the calorimeter clusters can be used. On top of this associating b-tagging information with the RC-jets is simple and requires only using the RC-jet's constituent sub-jets for which the b-tagging information is already available.

Reclustered jets are typically required to have a $p_T > 200$ GeV, $|\eta| < 2.0$ and $m > 50$ GeV although specific analyses may require tighter constraints. They also have the trimming procedure applied similarly to the LR-jets ($\Delta R = 0.2$ and $p_T/p_T^{RC} < 0.5$) but this has little effect since the subjets already have pile-up suppression applied as described in Section 7.3.

7.5.3 Jet Substructure

The clustering history of jets reconstructed with the anti-k_t algorithm is not particularly useful for deriving substructure variables. This is because the recombination sequence used will expand gradually through a soft subjet instead of the soft subjet being initially reconstructed before combination with the primary hard jet. Because of this the constituent clusters of the jets are also clustered with the k_t algorithm whose clustering history is then used to derive substructure variables with. Overall "substructure" is a loosely defined term and is used to describe variables that can help identify a jet using information from the constituents. They can be defined for both "big" ($\Delta R = 1.0$) and standard ($\Delta R = 0.4$) jets. However, typically they are more useful for identifying "big" jets as these can be used to reconstruct a wide variety of exotic high p_T candidates with collimated decays such as W, Z, Higgs bosons and top quarks as well as particles predicted by BSM theories. Standard jets would typically not encapsulate the decay products of more exotic particles e.g a Higgs unless the decay was ultra collimated (referred to as ultra boosted) requiring an extremely high p_T which would be extremely rare as it would only reconstruct candidates present in the very tail of the p_T distribution. Typically standard jets would represent jets originating from b-hadrons or lighter for which ATLAS produces dedicated tagging algorithms described in Section 7.4. Broadly speaking substructure variables can be separated into two categories: cluster based variables, these are derived directly from the calorimeter based cluster information; and non-cluster based variables which can either re-defined in terms of the constituent jets four-vectors (RC-jets) or use other information entirely such as b-tagging or ΔR.

Although many substructure variables exist the focus will be on ones useful for Higgs

($H \to \bar{b}b$) and top tagging which are the particles of interest for studies in Chapter 8. Classically substructure is used to refer to the cluster based variables : splitting-scale, N-subjettiness, energy correlation functions (ECF) and Q_w.

The variables $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ are defined from the splitting scales of the clustering algorithm [93]. In the case of an anti-k_t jet directly clustered from the calorimeter, such as the standard jet or LR-jet, the clustering history of the k_t algorithm is used as mentioned earlier and so are defined with $d_{ij} = min(p^2_{T,i}, p^2_{T,j})$ (see Section 7.3). They have units of mass with $\sqrt{d_{12}}$ referring to the final clustering step and $\sqrt{d_{23}}$ referring to the 2nd to last clustering step. In the case of RC-jets the input subjets are used to define the clustering history (but in k_t order) assuming the required subjet multiplicity is present (i.e >= 3 subjets for $\sqrt{d_{23}}$).

Q_w is another invariant mass quantity similar to the more general splitting scale variables. It's defined as the minimum invariant mass of a pair of subjets when decomposing an anti-k_t jet into 3 exclusive k_t jets. It is used to identify W boson decays inside of a "big" jet as would be the case for a top decay. For LR-jets it's defined from the calorimeter clustering information and for RC-jets it's defined directly from invariant masses of the subjets.

N-subjettiness (τ_N) [94] is used to measure "prong-like" structures in the calorimeter energy deposits that would result from i.e a collimated top decay (3 prong) or a collimated $H \to \bar{b}b$ decay. The τ_N series used to define N candidate subjets is:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\},$$

with k running over the constituent particles in a given jet, $p_{T,k}$ are their given transverse momenta and $\Delta R_{j,k} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is the distance in the rapidity-azimuthal plane between a candidate subjet j and a constituent particle k. The normalization factor $d_0 = \sum_k (p_{T,k} R_0)$ makes use of the original jet radius R_0. The τ_N variable can be interpreted as a measure of how well a given jet can be decomposed to N constituent k_t jets. The N-subjettiness variables commonly used are in-fact ratios of these with $\tau_{3,2} = \tau_3/\tau_2$ and $\tau_{2,1} = \tau_2/\tau_1$ having the best discrimination power for identifying 3 and 2 prong decays respectively. The defined ratios $\tau_{N,N_{-1}} = \tau_N/\tau_{N_{-1}}$ give a measure of how N prong vs N_{-1} prong-like a given jet is and take values $\tau_{N,N_{-1}} \in [0, 1]$. Using $\tau_{3,2}$ and top decays as an example we would expect the top-jet to have a lower value of $\tau_{3,2}$ compared to some competing 2 body decay.

Energy correlation functions (ECFs) are another class of substructure observables used to find prong-like structures in jets, unlike the N-subjettiness it however doesn't require a subjet finding procedure. Similarly to N-subjettiness it has a general form definable for N prongs. However, focus will be on the $D2$ variable which is defined as a ratio (like $\tau_{2,1}$)

and used to identify 2-prong like structures. It is derived from the calorimeter clustering information, for more information see [95].

Because the cluster based variables described above are typically derived directly from the calorimeter cluster information analyses need to consider how much MC statistics they require with how much extra discrimination/separation power these variables provide. This is because unless there are alternate ways to derive the variables, full detector simulation using Geant4 [40] would be required which is extremely CPU intensive. The alternate fast detector simulation uses AF2 [41] which makes use of the FastCaloSim [42] package for calorimeter simulation. This has a big impact as it is the most CPU intensive step reducing the simulation time from several minutes to a few seconds. However, the draw back of this is that only the four-vectors of jets are properly calibrated and the direct clustering history is poorly modelled and not usable. Because of this only Q_w and splitting-scales such as $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ as defined with the RC-jets are usable. For more information about detector simulation see Section 5.5.

7.6 Overlap Removal

To ensure that objects are reconstructed without any double counting of the calorimeter energy deposits / ID tracks overlap removal procedures are implemented to separate electrons (e), muons (μ) and jets. This is done based on the angular separation between two objects using $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The general procedure and order is:

1. If an e and μ share a track then the e is removed.
2. If a standard jet and e are within $\Delta R < 0.2$ of each other the jet is removed.
3. If a standard jet and e are within $0.2 \leq \Delta R < 0.4$ of each other the e is removed.
4. If a standard jet and μ are within $\Delta R < 0.4$ of each other and the jet has < 3 tracks the jet is removed.
5. If a standard jet and μ are within $\Delta R < 0.4$ of each other and the jet has ≥ 3 tracks the μ is removed.

The big-jets, either LR-jets or RC-jets, mentioned previously may also overlap with standard jets or leptons. However, this may be desirable i.e for ascribing b-tagging information to LR-jets etc. and is therefore not implemented ATLAS wide and left to the specific analyses to take into account. For the semi-leptonic $t\bar{t}H$ studies in sections 8.2 and 8.3 the RC/LR-jet is only used to describe the boosted Higgs and potential boosted hadronic top candidates with the single lepton (e / μ) coming from the leptonic top decay. Therefore for the RC/LR-jets an overlap removal requirement of $\Delta R > 1.0$ is required removing the RC/LR-jets in that case.

7.7 Missing Transverse Momentum

The ATLAS detector is not able to detect every particle that passes through it, with neutrinos ν being a frequent example as they pass through the detector without any interactions. There is also the possibility that some exotic particle, predicted by BSM theories, to pass through undetected. Because of this the missing momentum from an event may be used as a proxy to gauge what might have been present in a given event. However, due to to the exact proton momentum fraction of the partons in the hard scattering event being unknown the expected momentum sum of the produced particles is not known. This is because only the centre of mass energy of the protons is known. Despite this it is expected that the transverse momentum to the beam line would sum to 0 assuming that the partons collide head on in the z direction. It is then assumed that if the transverse momentum in an event doesn't sum to 0 that it's the result of an undetected particle and is referred to as "missing transverse momentum" (p_T^{miss}) and is a vector with magnitude "missing transverse energy" (E_T^{miss}). This could be cause by both a mismeasurement of detected particles or from some undetected particle.

The measurement of p_T^{miss} is difficult because it requires every detector subsystem and requires a complete unambiguous representation of the hard scattering. It is reconstructed by summing the momentum transverse to the beam-line of every reconstructed object post overlap removal which includes: leptons, photons, standard jets as well as any tracks associated with the primary vertex that could originate from soft scattering.

Uncertainties on P_T^{miss} come from propagating the scale and resolution uncertainties from the reconstructed objects used to determine it including additional scale and resolution uncertainties associated with the soft terms. Because the semi-leptonic $t\bar{t}H$ analyses in sections 8.2 and 8.3 don't use P_T^{miss} for event selection and only use it to aid in event reconstruction, as BDT inputs, these uncertainties have a minimal impact on them.

Chapter 8

$t\bar{t}H$ Analyses

This chapter focuses on the $t\bar{t}H$ cross-section analyses with the $H \to b\bar{b}$ decay mode (see Figure 8.1). It is split into 3 sections: In Section 8.1 a boosted all-hadronic feasibility study is carried out to see if a region can be found with enough statistical significance for signal sensitivity. In Section 8.2 an overview of the lepton + jets analysis [1] results using the full Run-2 ATLAS data set is given, which is a combination of resolved di-lepton, resolved semi-leptonic and boosted semi-leptonic channels. This is accompanied by a more detailed explanation of the boosted region and its analysis techniques. In Section 8.3 the ongoing studies for the boosted region of lepton + jets legacy analysis are described. The Legacy analysis is a re-analysis of the full Run-2 ATLAS data set with updated techniques.

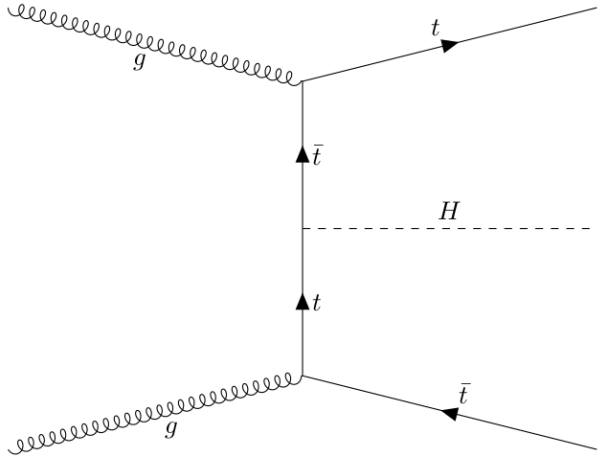


Figure 8.1: Example of a tree-level Feynman diagram of the $t\bar{t}H$ t-channel process.

When analysing $t\bar{t}H$ with $H \to b\bar{b}$, events can be further split into 3 distinct decay channels depending on the top decaying hadronically or leptonically (see Figure 8.1) into: all-hadronic where both tops decay hadronically giving an 8-jet 4-b final state, semi-leptonically where only one of the tops decay leptonically giving a 6-jet 4-b final state and di-lepton where both of the tops decay leptonically giving a 4-jet 4-b final state. The primary challenge in these channels is separating the $t\bar{t}H$ signal from the primary background

$t\bar{t}$+b-jets which, when produced in conjunction with a pair of b-quarks can reproduce the final states mentioned above. Gluon induced backgrounds can be differentiated either through their spins, as gluons are spin 1 while the Higgs is spin 0, or via their colour charges as gluons carry $C = 2$ while the Higgs has no colour charge.

There is a further separation to be made based on the p_T of the jets, as decay products with a high p_T will have collimated decay products (see Figure 8.2). Because of this substantially different analysis techniques can be employed compared to resolved decays such as clustering all the decay products into one big jet with $\Delta R = 1$ either with LR-jets (see Subsection 7.5.1) or RC-jets (see Subsection 7.5.2). Since all the objects decay products are encapsulated into one big jet another tool that can be used in jet identification are sub-structure variables (see Subsection 7.5.3). Additionally the high p_T regime is more sensitive to BSM physics which is an additional motivation for boosted (high p_T) studies.

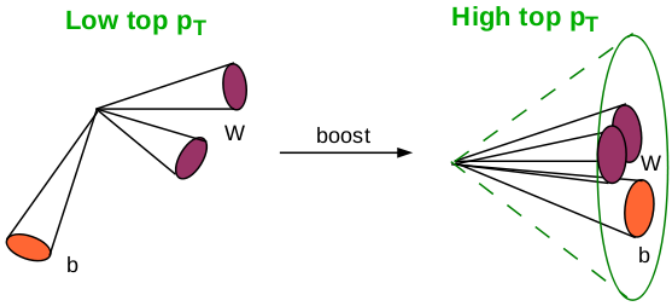


Figure 8.2: Resolved vs boosted top decays.

8.1 All hadronic feasibility studies

The aim of this study is to assess the feasibility of an all-hadronic $t\bar{t}H$ analysis in the boosted region. It was conducted using the POWHEG + Pythia 8 [96–101] $t\bar{t}H$ and $t\bar{t}$ mc16a MC samples matching a luminosity of $36fb^{-1}$. This mirrors the nominal samples used in the Lepton + Jets Analysis [1]. An all-hadronic $t\bar{t}H$ decay (see Figure 8.3) has an 8 jet final state of which 4 jets originate from b decays. The 2 b-jets originating from the Higgs decay will provide the signature for identification so correctly separating b-jets originating from Higgs and top decays is important. Because of this identifying the b-jets is paramount for the analysis and the MV2c10 b-tagging algorithm developed centrally by ATLAS was used for these studies, the performance of which can be seen in Sections 7.4.1.

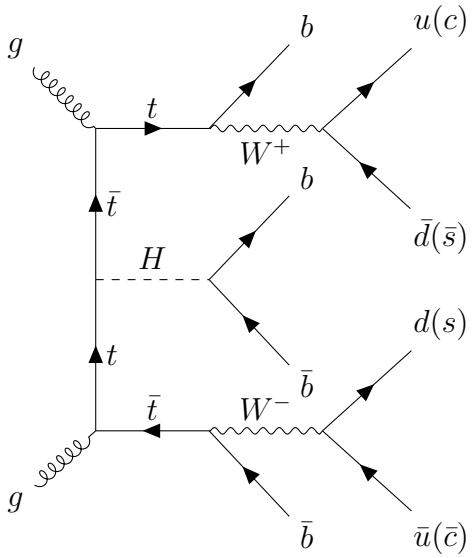


Figure 8.3: Feynman diagram of $t\bar{t}H$ decaying all-hadronically. This is the most common final state as top quarks decay to b-jets $sim100\%$ while W bosons decay to u, d or c, s $\sim 50/50$.

The primary challenge in this analysis is to separate the signal from the background processes with $t\bar{t}$ + jets being the largest known background that can be modelled. From Figure 8.4 we can see that $t\bar{t} + b\bar{b}$ can have the exact same final state (8 jet 4 b) as our signal. There is also a large QCD-multijet background which can't be modelled well with MC.

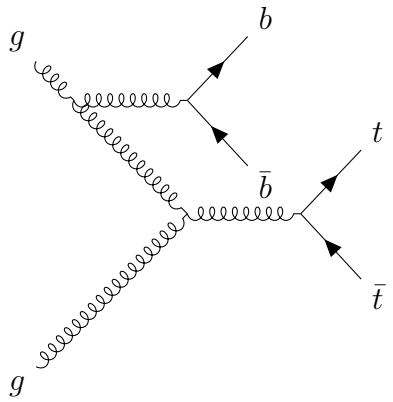


Figure 8.4: Feynman diagram of $t\bar{t} + b\bar{b}$.

To gauge the feasibility of this study the statistical significance $Z \approx s/\sqrt{b}$ is used. This approximation is valid when $s << b$ (see Section 6.1). Because of the large unknown QCD-multijet background two metrics are used $s/\sqrt{b}$ where b is approximated by $t\bar{t}$ + jets, the dominant background, and $data/b$ which is used to gauge how much of the QCD-multijet background exists in a given region. With these two metrics sets of cuts are proposed which are used to define regions in order to maximise $s/\sqrt{b}$ and minimise $data/b$.

A purely boosted strategy is used where both Higgs and top candidates are modelled with boosted objects, in this case RC-jets (see Section 7.5.2) with Higgs and top tags initially defined through cuts on the subjets b-tags:

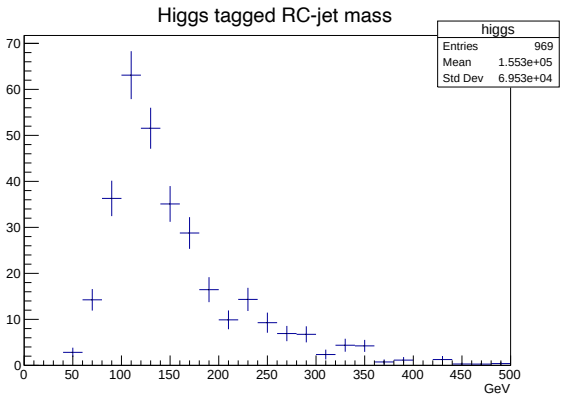
- top tag requires an RC-jet with $p_T > 250$ GeV, containing 1 b-tagged and 1 untagged subjet.
- Higgs tag requires an RC-jet with $p_T > 200$ GeV, containing $\geq 2b$-tagged subjets.

and later with a dedicated boosted Higgs DNN tagger, which was originally developed for the lepton + jets analysis. The DNN tagger makes use of a fair amount of substructure variables including the subjets b-tagging information among others, see Section 8.2 for more details.

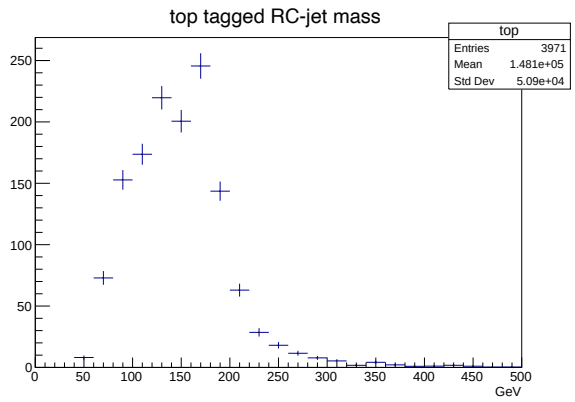
The initial selection, referred to as the pre-selection, made using the 85% b-tagging WP is:

- 0 leptons with $p_T > 10$ GeV
- 4 jets $p_T > 25$ GeV
- $\geq 2b$-tags
- ≥ 1 RC-jet with $p_T > 200$ GeV
- 1 top tag.

The mass distributions of the Higgs and top candidates in this region are shown in Figure 8.5. From Figure 8.5a we can see two peaks: the expected one around 120 GeV, along with an unexpected peak around 225 GeV.



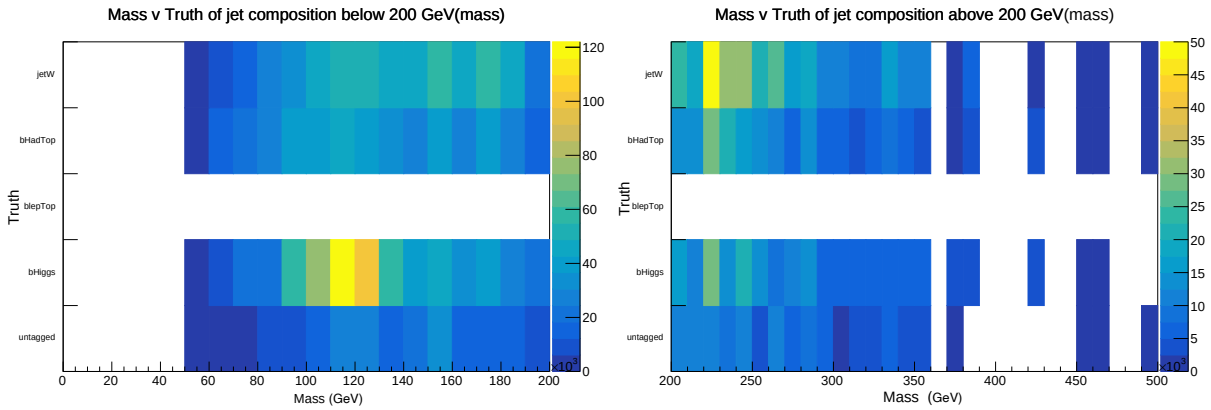
(a) RC-jet mass distribution of a Higgs candidate.



(b) RC-jet mass distribution of a top candidate.

Figure 8.5: Invariant mass distributions of boosted Higgs / top candidates using the cut based definitions made with a POWHEG + Pythia 8 $t\bar{t}H$ MC sample using the initial preselection mentioned above.

This secondary anomalous peak was investigated using the truth information with which the MC samples were generated. To that end the samples were split at 200 GeV with truth information given for each bin as can be seen in Figure 8.6. From Figure 8.6a we can see that the expected peak around 120 GeV does indeed come from an excess of Higgs decays and from Figure 8.6b we can see that the anomalous mass peak at 220 GeV comes from an excess of W decays which are being mistagged by our Higgs tag requirement of an RC-jet with $\geq 2b$-tagged subjets.

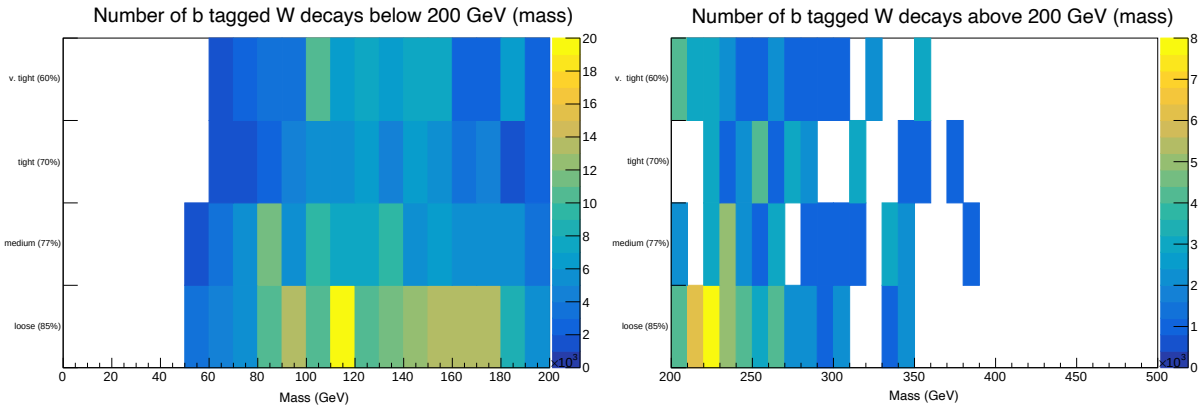


(a) Truth information for mass ranges $[0, 200)$ GeV

(b) Truth information for mass ranges $[200, 500)$ GeV

Figure 8.6: Truth information for Higgs tagged RC-jets per invariant mass bin.

Since the W boson decays hadronically 67.4% of the time, with a cX final sate fraction of 33.3% [13], it can be assumed that the mistagged W decays are due to c / *light* flavoured jets being mistagged as b-jets by the MV2c10 algorithm. To gauge this effect the b-tagging scores of these mistagged W jets is looked at in Figure 8.7 which does indeed show this to be the case. However, it can also be seen that this effect can be mitigated by using a tighter b-tagging WP due to the W decays being mistagged primarily at the 85% WP where the MV2c10 algorithm suffers in c and *light* jet rejection (see Section 7.4.1 for MV2c10 performance).



(a) b-tagging information for mass ranges $[0, 200)$ GeV

(b) b-tagging information for mass ranges $[200, 500)$ GeV

Figure 8.7: b-tagging truth information for RC-jets origination from true W decays.

With the composition of the cut based Higgs tags understood attention was then focused on defining a signal region with a high $s/\sqrt{b}$ and low $data/b$. To that end the jet multiplicity of both $t\bar{t}H$ and $t\bar{t}$ + jets matching the MC16a campaign (see Chapter 5) was looked at as well as the 2015/16 data ($36fb^{-1}$), this can be seen in Figure 8.8. The $t\bar{t}$ + jets background was split by jet flavour, into $t\bar{t}$ + b, $t\bar{t}$ + c and $t\bar{t}$ + $light$, to see how these respond to b-tagging and jet multiplicity.

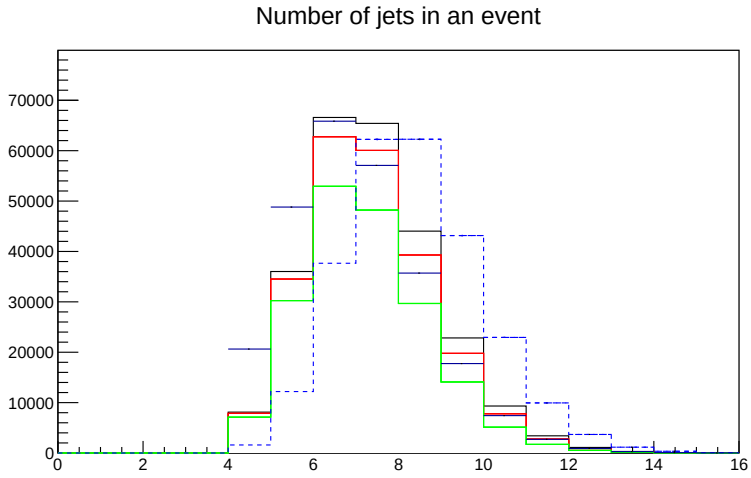


Figure 8.8: Jet number in events of signal ($t\bar{t}H$), stacked backgrounds ($t\bar{t}$ + b, $t\bar{t}$ + c, $t\bar{t}$ + light) and 2015/16 data at $36fb^{-1}$ (marked with dark blue error bars). The signal and data have both been scaled to the background for shape comparisons. All integral values are calculated to match data with a luminosity of $36fb^{-1}$.

From Figure 8.8 a cut ≥ 7 jets was proposed along with cuts ≥ 1 Higgs tagged RC-jet and ≥ 4 b-tagged jets. These were implemented using both the 85% and 60% b-tagging WP. The event yields after each cut can be seen in Table 8.1 and Table 8.2. From Table 8.1 we can see that the full set of cuts provide an initial region with a decent significance

($s/\sqrt{b} = 2.00$, $s/b = 0.022$) when only considering the $t\bar{t} + jets$ background. However the estimate of the QCD-multijet background $data/b = 5.24$ implies that although it has been more than halved from the initial rate present at the pre-selection the true significance of the current selection would be insignificant. From Table 8.2 we can see that tightening the b-tagging WP from 85% to 60% did not improve things giving a final significance ($s/\sqrt{b} = 1.88$, $s/b = 0.126$). We can see that while it did constrain the $t\bar{t}$ + c and $t\bar{t}$ + $light$ fractions of the $t\bar{t}$ + $jets$ background, as expected, it did not constrain the $data$ to the same degree implying that large fractions of the QCD-multijet background is made up of true b-hadrons. This is highlighted by the effect of the cut requiring 4-b tagged jets.

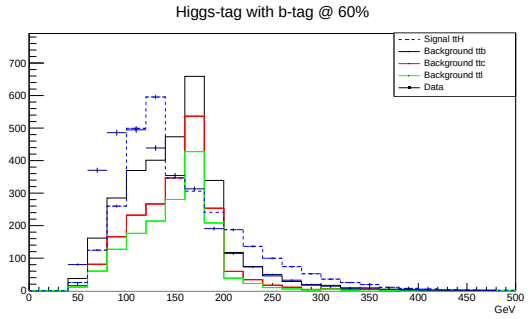
	$t\bar{t}H$	$t\bar{t}+b$	$t\bar{t}+c$	$t\bar{t}+l$	$t\bar{t}$ (all)	data	data/b	$s/\sqrt{b}$
Pre-selection	1000	21200	46200	190000	257000	2950000	11.5	1.98
Higgs-tag req	337	5010	9360	38200	52600	256000	4.86	1.47
7-Jet req	271	3750	5910	16900	26600	137000	5.17	1.67
4 b-tags req	184	2360	2410	3650	8430	44100	5.24	2.00

Table 8.1: Cut flow table of initial cuts with b-tagging at 85% rounded to 3 s.f.

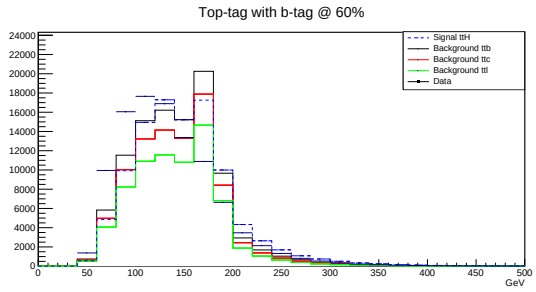
	$t\bar{t}H$	$t\bar{t}+b$	$t\bar{t}+c$	$t\bar{t}+l$	$t\bar{t}$ (all)	data	data/b	$s/\sqrt{b}$
Pre-selection	585	12200	16200	71800	100000	465077	4.64	1.84
Higgs-tag req	89.0	1020	446	1590	3060	18500	6.04	1.61
7-Jet req	71.2	769	264	661	1690	9800	5.78	1.73
4 b-tags req	27.9	205	10.3	5.97	221	1430	6.45	1.88

Table 8.2: Cut flow table of initial cuts with b-tagging at 60% rounded to 3 s.f.

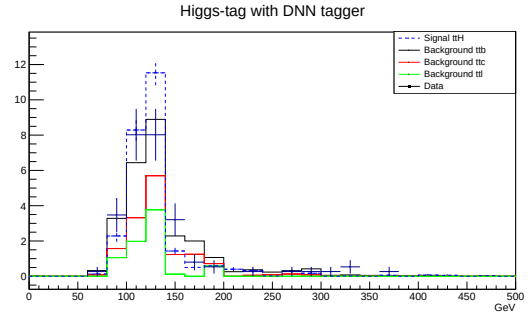
Following these results, the Higgs and top tags were looked at with comparisons made between cut based versions and a DNN developed for the lepton + jets boosted region, which is described in Section 8.2. The shape distributions of Higgs / top tagged RC-jets for both methods can be seen in Figures 8.9. The DNN was trained with the RC-jets invariant mass as well as making use of substructure information, such as the subjet multiplicity and the subjets b-tagging scores, among others. It can be seen that Higgs / tops selected with the DNN exhibit mass sculpted peaks around both the Higgs and top mass when compared to the cut based methods which have broader distributions. A new set of cuts was proposed selecting ≥ 2 top tags, ≥ 1 Higgs tag and imposing mass requirements on the top tagged RC-jets to be within 50 GeV of the top mass ($|M_{jet} - M_{top}| < 50$ GeV), the results of which can be seen in Tables 8.3 and 8.4.



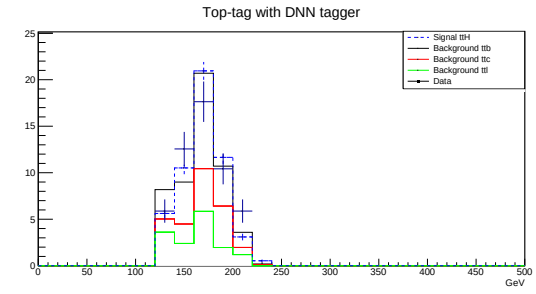
(a) Higgs tagged RC-jets using cut based method.



(b) Top tagged RC-jets using cut based method.



(c) Higgs tagged RC-jets using DNN.



(d) Top tagged RC-jets using DNN.

Figure 8.9: Comparison of DNN vs cut based methods for Higgs / top identification. The signal and data have both been scaled to the background for shape comparisons. All integral values are calculated to match data with luminosity of $36fb^{-1}$.

cuts	$t\bar{t}H$	$t\bar{t}b$	$t\bar{t}c$	$t\bar{t}l$	$t\bar{t}$(all)	data	data/b	$s/\sqrt{b}$
Pre-selection	585	12200	16200	71800	100000	465000	4.64	1.85
≥ 2 top-tag	68.3	1270	306	447	2020	8020	3.97	1.52
≥ 1 Higgs-tag	1.78	12.06	0.56	0.00	12.6	70	5.55	0.50
≥ 1 top-tag with mass req	1.54	9.71	0.32	0	10.0	51	5.08	0.49
≥ 2 top-tags with mass req	0.68	3.88	0.11	0	4	15	3.75	0.34

Table 8.3: Cut flow table of second iteration of cuts with cut based Higgs / top tagging using the 60% b-tagging WP rounded to 3 s.f.

cuts	$t\bar{t}H$	$t\bar{t}b$	$t\bar{t}c$	$t\bar{t}l$	$t\bar{t}$(all)	data	data/b	$s/\sqrt{b}$
Pre-selection (DNN tagger)	31.4	559	1560	7830	9950	25500	2.56	0.31
≥ 2 top-tag	31.4	559	1560	7830	9950	25500	2.56	0.31
≥ 1 Higgs-tag	1.5	12	6.7	7.5	26.2	98	3.74	0.29
≥ 1 top-tag with mass req	1.5	12	6.7	7.5	26.2	98	3.74	0.29
≥ 2 top-tags with mass req	1.5	12	6.7	7.5	26.2	98	3.74	0.29

Table 8.4: Cut flow table of second iteration of cuts with DNN Higgs / top tagging rounded to 3 s.f.

From Table 8.3, we can see that the significance drops dramatically giving an $s/\sqrt{b} = 0.34$, $s/b = 0.17$ unfortunately $data/b = 3.75$ is still high and so the region would not provide signal sensitivity. The story is much the same when using the DNN for Higgs / top tagging as can be seen in Table 8.4 with $s/\sqrt{b} = 0.29$, $s/b = 0.057$ and $data/b = 3.74$. The most notable differences from the cut based tagging method are that the top mass requirements have no effect, due to the mass sculpting by the DNN (see Figures 8.9c and 8.9d), and that the second top tag requirement (with the first being in the pre-selection) has no effect, which implies a certain consistency in its tagging. Meaning that in all the events that the DNN could identify one top in, it could also identify a second top which is to be expected given that it's a $t\bar{t}H$ sample. Graphs of the cut-flow Tables 8.1, 8.2, 8.3 and 8.4 can be found in Appendix B providing visual aides compared to the tables above.

Overall it was decided that an all-hadronic boosted analysis seems infeasible with current techniques as the irreducible QCD-multijet background combined with the already challenging signal like background process provided by $t\bar{t}$ + jets yields little sensitivity and is extremely difficult to separate from signal. Most likely data driven methods would be needed such as the tag rate efficiency (TRF) method employed in [102] (a previous resolved all-hadronic $t\bar{t}H$ analysis) to estimate the QCD-multijet background as well as using a hybrid approach of resolved and boosted topologies with the hope of gaining some performance from the extra boosted events.

8.2 Lepton + Jets Full Run-2 Analysis

This section focuses on my contribution to the boosted region of the Lepton + Jets Full Run-2 Analysis. It is a measurement of the $t\bar{t}H$ production cross-section in the $H \to b\bar{b}$ channel using the full Run-2 ATLAS data set of $139fb^{-1}$ at $\sqrt{s} = 13$ TeV [1]. Initially an overview of the analysis is given with subsections: 8.2.1, 8.2.2 and 8.2.3 focusing on the boosted region. In Section 8.2.1 the boosted signal region is described. In Section 8.2.2 the boosted Higgs classification DNNs input features pre- and post-fit modelling is looked at in a validation region. Finally in Section 8.2.3 the boosted signal regions contribution to the combined fit result is shown. For full details of the analysis please see the original paper [1].

The analysis makes use of semi-leptonic (see Figure 8.10) or dilepton top decays which constrain the QCD-multijet background present in the all-hadronic decay channel. The measurement is a combination of three separate regions: resolved dilepton, resolved single-lepton and boosted single-lepton with events split between them based on the number of leptons, the number of jets and the number of b-tagged jets (b-jets). Leptons (e / μ) are reconstructed as ATLAS objects and are described in detail in Section 7.2, EMtopo jet collections are used described in Subsection 7.3.1, b-jets are tagged with the MV2c10

b-tagging algorithm described in Subsection 7.4.1 and boosted objects are modelled with RC-jets described in Subsection 7.5.2. Events are further split into signal-rich and signal-depleted regions, dubbed signal regions (SR) and control regions (CR) respectively, with machine learning algorithms and are combined in a profile likelihood fit (see Section 6.1). The outputs of these machine learning algorithms are used as the primary discriminant for signal extraction and enter into the fit, which simultaneously determines the event yields for signal and important background processes, while constraining the overall background model with the given systematic uncertainties. Additionally the measurement is also repeated by recombining events from all regions based on reconstructed Higgs p_T, into bins defined by the simplified template cross-section (STXS) formalism [37], with the aim of separating the measurement and interpretation steps in order to reduce theory dependencies being folded into measurements. Truth information from the MC events, prior to their Higgs decay, is used to calculate the cross-section for each of the STXS bins, defined to minimise correlations with different production processes between bins.

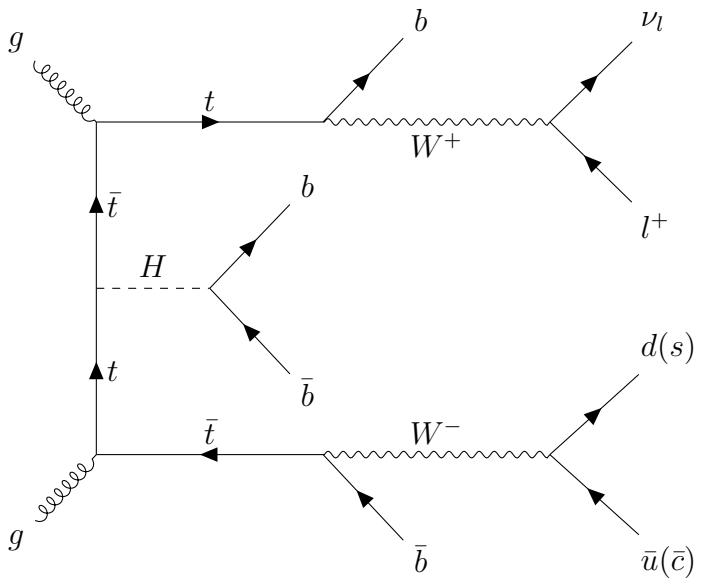


Figure 8.10: $t\bar{t}H$ t-channel process, decaying semi-leptonically. One top quark decays leptonically while the other decays hadronically.

After lepton reconstruction described in Section 7.2 they are also required to satisfy additional identification and isolation criteria to increase background rejection: e / μ must pass the Tight / Medium identification criteria and the $Gradient$ / $FixedCutTightTrackOnly$ isolation criteria respectively [103, 104].

The dilepton region requires events with exactly 2 leptons with opposite electric charge, with at least one lepton having $p_T \geq 27$ GeV and match a corresponding object at trigger level. In the ee and $\mu\mu$ channels, the leptons combined invariant mass must be > 15 GeV and outside the Z-bosons mass window of $83 - 99$ GeV. In order to maintain orthogonality with other $t\bar{t}H$ channels [105], events are vetoed if they contain one or more τ_{had} candidates.

It has three CR defined, with the loosest criteria requiring only ≥ 3 jets and ≥ 3 b-jets. The SR for the dilepton region requires ≥ 4 jets and ≥ 4 b-jets with Higgs candidates being reconstructed with the same BDT used in the previous analysis [23]. The classification BDT (used to identify $t\bar{t}H$ events) has been re-trained with the signal and components of the nominal background model presented in the analysis and is trained only against $t\bar{t}+b\bar{b}$ events. The SR is further split into the STXS bins of: $[0, 120)$, $[120, 200)$, $[200, 300)$, $[300, \infty)$ GeV where the final two STXS bins $[300, 450)$ and $[450, \infty)$) have been merged due to the low number of events expected to be present. For the full details of the SR and CR see Table 8.5 or the flow chart Figure C.1 in Appendix C.

The single-lepton region requires events to have exactly 1 lepton with $p_T > 27$ GeV and match a corresponding object at trigger level with events being vetoed if they contain two or more τ_{had} candidates to maintain orthogonality with other $t\bar{t}H$ channels as before. Events are required to have ≥ 4 jets and ≥ 4 b-jets at the 85% WP. The region is split into boosted and resolved categories with events being assigned to the boosted category if a boosted Higgs candidate is found as well as containing 2 extra b-jets at the 77% WP not contained within the RC-jet of the boosted Higgs candidate. A boosted Higgs candidate is primarily identified through the use of a classifier DNN (see Subsection 8.2.1), filling the same role as the reconstruction BDTs in the resolved regions. It requires an RC-jet with $p_T > 300$ GeV, have $m \in [100, 140]$ GeV, a DNN score $P(H) > 0.6$ and exactly two b-jets tagged at the 85% WP. Events are then further further classified into the boosted signal region with a classification BDT score ≥ -0.05 which aims to reconstruct $t\bar{t}H$ events. The boosted classification BDT is trained against all available background processes. The boosted SR gets further split into the $[200, 300)$, $[300, \infty)$ STXS bins. Events with a boosted BDT score < -0.05 get pushed back into the resolved single-lepton category. For the resolved single-lepton category the SR further requires ≥ 6 jets with ≥ 4 b-jets tagged at the 70% WP. The resolved single-lepton region similar to the resolved dilepton region also re-uses the Higgs reconstruction BDT from the previous analysis and has its classification BDT trained against $t\bar{t}$ + jets background. The resolved single-lepton SR is again further split into the STXS bins of: $[0, 120)$, $[120, 200)$, $[200, 300)$, $[300, 450)$ and $[450, \infty)$. For the full details of the SR and CR see Table 8.5 or the flow chart Figure C.1 in Appendix C.

The nominal signal MC sample is modelled with Powheg + Pythia 8 [96–101] in the 5 flavour scheme (FS) with all Higgs decay modes considered. Here Powheg provides the matrix element calculation and Pythia 8 is used to shower the event simulating the "pile-up type" interactions originating from soft scattering events. The primary background $t\bar{t}$ + jets is again modelled in the 5FS using Powheg + Pythia 8. The $t\bar{t}$ + jets is further split based on jet type into the b, c and *light* flavours. However, due to the 5FS vs 4FS modelling uncertainty arising in the case of $t\bar{t}$ + $\geq b$ various other samples are also

Region	Dilepton				Single-lepton			
	$SR^{\geq 4j}_{\geq 4b}$	$CR^{\geq 4j}_{3b\ \mathrm{hi}}$	$CR^{\geq 4j}_{3b\ \mathrm{lo}}$	$CR^{3j}_{3b\ \mathrm{hi}}$	$SR^{\geq 6j}_{\geq 4b}$	$CR^{5j}_{\geq 4b\ \mathrm{hi}}$	$CR^{5j}_{\geq 4b\ \mathrm{lo}}$	SR_{boosted}
#leptons	= 2				= 1			
#jets	≥ 4			= 3	≥ 6	= 5		≥ 4
#b-tag @85%	–				≥ 4			
#b-tag @77%	–				–			≥ 2†
#b-tag @70%	≥ 4	= 3			≥ 4			–
#b-tag @60%	–	= 3	< 3	= 3	–	≥ 4	< 4	–
#boosted cand.	–				0			≥ 1
Fit input	BDT	Yield			BDT/Yield	$\Delta R^{\mathrm{avg}}_{bb}$		BDT

Table 8.5: Analysis region definition split by number of leptons, jets, b-jets with SRs further split into STXS bins based on the reconstructed Higgs p_T. In the boosted SR b-jets with † means that those jets are not contained within the boosted Higgs candidate RC-jet [1].

generated in order to apply a systematic uncertainty for this. For a full list of samples used see Table C.1 Appendix C.

Overall the analysis strategy used in each region is broadly similar and makes use of machine learning techniques in two main steps: reconstructing / identifying the Higgs boson candidate, and classifying the $t\bar{t}H$ signal events.

For both resolved regions BDTs are employed for Higgs reconstruction, they are re-used from the previous analysis [23] and match the reconstructed jets to partons emitted from top-quark and Higgs boson decays. It reconstructs W, top and Higgs candidates from combinations of jets and leptons making use of the b-tagging information to do so. The combinations of jets with the highest BDT scores are assigned to these objects thus allowing for the computation of the Higgs candidates kinematics (such as p_T) by summing the four momenta of the two jets assigned to it. For the boosted region a DNN is employed for Higgs identification described in detail in the next Section 8.2.1.

The $t\bar{t}H$ event classification BDTs for the resolved channels are trained with the kinematic variables of pairs of jets and leptons (m, ΔR separation), the reconstruction BDT outputs as well as their b-tagging information for jets. In addition for the resolved single-lepton channel a likelihood discriminant method describing the signal and background probabilities of the jet combinatorics is also added. The boosted region also uses a BDT for event classification which makes use of the DNNs output as before, again this will be described in detail in the next Section 8.2.1.

The rest of this chapter will focus on the boosted region where the full Higgs identi-

fication DNN will be described, as well as the boosted $t\bar{t}H$ event classification BDT. On top of this the input variables to the DNN are also studied in a 1 top 0 Higgs validation region to look at the Data/MC agreement of all of the input features it makes use of. A summary of the results from all regions is then provided highlighting the major systematic pulls in the final fit.

8.2.1 Boosted Region

As described before the boosted single-lepton region requires 1 lepton, 4 jets and 4 b-jets at the 85% WP, with events being split between the boosted and resolved regions on the requirements that: a boosted Higgs boson candidate is found, ≥ 2 additional b-jets not contained within the RC-jet of the Higgs candidate and that it has a boosted BDT score ≥ -0.05. It contributes to two STXS bins with reconstructed Higgs $p_T \in [300, 450) \cup [450, \infty)$.

Boosted objects are modelled with RC-jets (see Subsection 7.5.2) with $p_T \geq 300$ GeV, $m \geq 50$ GeV and $\eta \leq 2$. Events with electrons found to be within $\Delta R \geq 1$ of an RC-jet are removed so there is no overlap with RC-jets. Higgs candidates are modelled with RC-jets and are required to have: $100 \leq m < 140$ GeV, contain exactly 2 b-tagged jets tagged at the 85% WP and require a DNN score $P(H) \geq 0.6$ which is in effect the defining criteria as it implicitly, although not strictly, requires the above conditions. The DNN cut of $P(H)$ was chosen to maximise both $\frac{s}{\sqrt{b}}$ and selection efficiency ($\frac{\text{sel. } t\bar{t}H}{\text{tot. } t\bar{t}H}$) defined with the MC truth information. If more than one Higgs candidate is found the one with the closest mass to the Higgs mass is selected.

The classification DNN is trained to classify RC-jets into the three categories of: Higgs, top and QCD and is the primary algorithm used to identify Higgs candidates. It was trained with the Keras python package using tensorflow as a backend [63, 64]. The DNN consists of 3 simply connected layers each with a width of 100 neurons, uses the "ReLU" activation and has a softmax final layer. The softmax final layer ensures that the total output probabilities $P(H) + P(t) + P(QCD) = 1$ where H and t correspond to the Higgs and top categories. It uses categorical cross-entropy as the loss function and stochastic gradient decent (SGD) as the optimiser. It was trained with 17 input features with all jets being p_T ordered in the event:

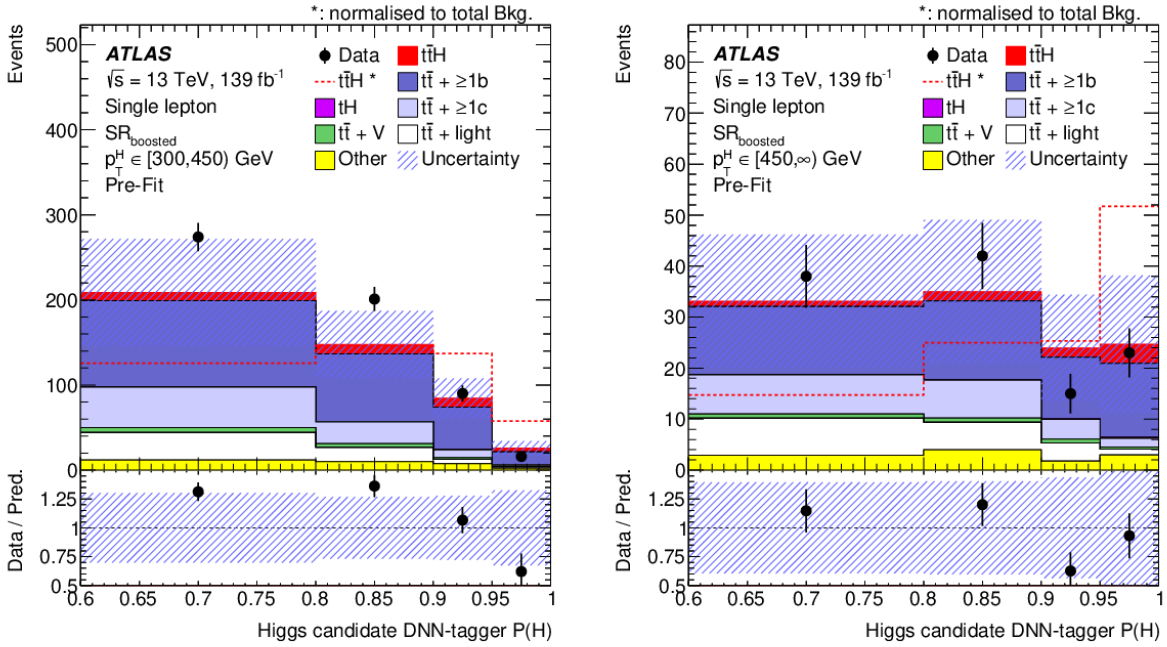
1. the invariant mass of the RC-jet,
2. the number of subjets in the RC-jet,
3. the b-tagging WP of the leading subjet,
4. the b-tagging WP of the sub-leading subjet,

5. the maximum b-tagging WP of all subjets contained within an RC-jet,
6. the minimum b-tagging WP of all subjets contained within an RC-jet,
7. the remaining sum of b-tagging WP of subjets minus the WPs of the leading and sub-leading subjets,
8. the p_T of the leading subjet,
9. the p_T of the sub-leading subjet,
10. the substructure variable d_{12},
11. the substructure variable d_{23},
12. the substructure variable Q_w,
13. the invariant mass of all b-tagged subjets at the loosest 85% WP,
14. the invariant mass of all non b-tagged subjets,
15. the ΔR separation of the leading and sub-leading subjets,
16. the maximum ΔR separation between subjets in the RC-jet,
17. the minimum ΔR separation between subjets in the RC-jet.

As can be seen from the above list, it makes use of primarily RC-jet substructure type information (see Section 7.5.3) with the only kinematic feature employed being the invariant mass. Groupings are made of the most useful discriminators such as b-tagging and ΔR separation such that they maximise the separation power between the classes (Higgs, top and QCD), this can be seen in identical studies using the new jet collection in Subsection 8.3.1 for the Legacy Analysis. Classically when deep learning techniques are used these sort of groupings to create high separation features are not used because it is hoped that the DNNs can generalize these sort of transformations from the data by themselves. While this may be a possibility it tends to require a huge amount of data which due to the limited training data available in this analysis is not a possibility, instead this approach is taken which helps both increase the speed of convergence during training to the "optimal" minima and the performance of the trained network achievable with the given training data set. To further help maximise the use of the training MC available, which is re-used from the analysis, it is trained on sub-sets of the samples split into odd and even categories based on event number with the model trained on odd being applied to even and vice versa. It is important to not arbitrarily split the MC samples because events originating from different MC campaigns will have slightly different normalisations

in order to account for the radiation damage sustained by the detector over time. Splitting the data this way allows for an even distribution of events from each campaign to be seen by the network during training.

The pre-fit $P(H)$ distribution can be seen below for the boosted Higgs classification DNN. Then its performance is compared to the Higgs classification BDTs used in the resolved regions. This is achieved with confusion matrices, assessing the migration between the truth and reconstructed Higgs-p_T bins. Finally the pre-fit Data/MC agreements of the SR reconstructed Higgs p_T is shown.



(a) Data/MC agreement of the $P(H)$ DNN output in the $p_T \in [300, 450)$ GeV.

(b) Data/MC agreement of the $P(H)$ DNN output in the $p_T \in [450, \infty)$ GeV.

Figure 8.11: Pre fit distributions of Data/MC agreement of the DNN Higgs tagger $P(H)$ distribution in the boosted SR STXS bins. The dashed red line shows the signal scaled to the total background used to highlight shape preferences between the signal and background. The uncertainty band includes all uncertainties and correlations [1].

From Figure 8.11 we can see that there are no obvious shape disagreements in the Data/MC agreement with all bins staying within the uncertainty bands. Interestingly $t\bar{t}H$ events, as can be seen from the dashed red line, in the final bins of DNN $P(H)$ behave quite differently between the $[300, 450)$ and $[450, \infty)$ GeV p_T ranges.

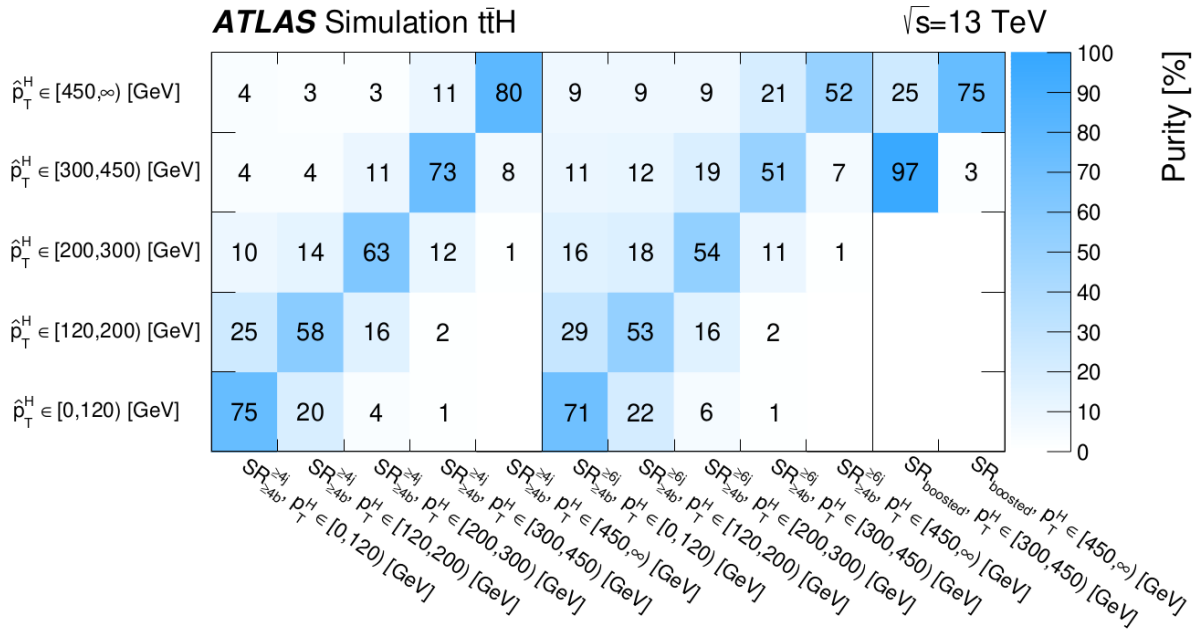
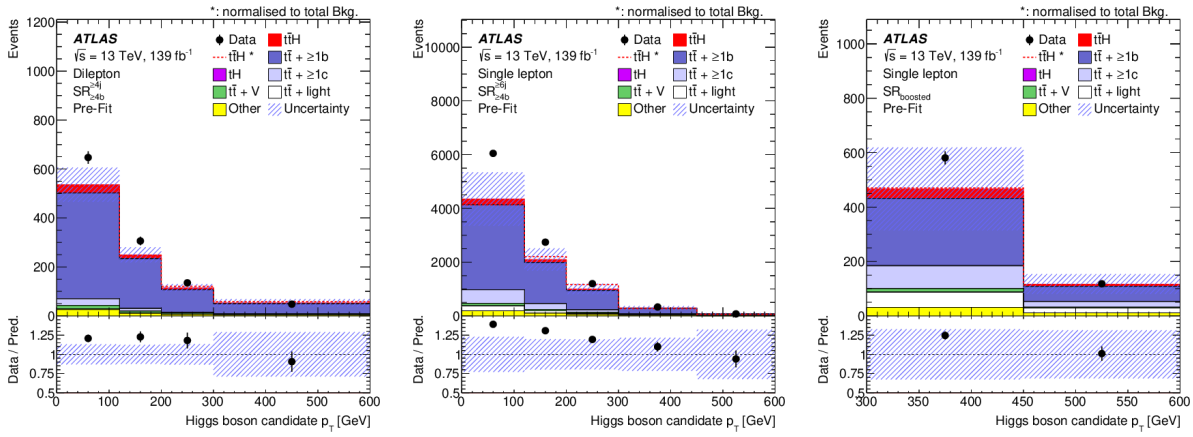


Figure 8.12: Performance comparison of the Higgs reconstruction algorithms for each STXS bin split for each region using confusion matrices. Each row of "truth" $\hat{p}_T^H$, the matrix shows the percentage of all Higgs boson candidates with reconstructed p_T in the various bins of each region with left for dilepton, middle for resolved single-lepton and right boosted single-lepton [1].

From Figure 8.12 we can see that compared to the resolved reconstruction BDTs the boosted reconstruction DNN works exceptionally well albeit with a relatively large, but expected, migration from the $[450, \infty)$ truth bin to the $[300, 450)$ reconstruction bin which implies that not all the decay products were captured by the RC-jet.



(a) Dilepton SR. (b) Semileptonic resolved SR. (c) Semileptonic boosted SR.

Figure 8.13: Pre fit Data/MC agreements of the various SR reconstructed Higgs p_T [1].

From Figure 8.13 we can see that there is some miss-modelling present in the Data/MC of the reconstructed Higgs p_T in each SR. This is corrected for collectively in each region in the final fit and shows up as a large pull on the p_T^{bb} shape uncertainty in Figure 8.30.

Each region then uses a BDT to classify signal events and separate them from backgrounds. With the boosted specific BDTs most important inputs being: the DNN output $P(H)$ for the Higgs candidate; the sum of the b-tagging WP for each candidate (Higgs, hadronic top and leptonic top); the hadronic tops invariant mass; jet multiplicity and the fraction between the sum of b-tagging WP excluding the Higgs and hadronic top candidates over the total sum b-tagging WP.

Tops are reconstructed in a fall through scheme, with the preferred hadronic top candidate being reconstructed with an RC-jet that has been tagged using the DNN $P(t)$ score. In all cases candidates with the closest mass to the top mass will be selected from the event. If no RC-jet candidate is found for the hadronic top then combinations of jets are used along with the leptonic top, with the leptonic top taking preference and selecting the combinations with the closest mass to the top mass. The leptonic top is reconstructed by selecting the lepton in the event together with the ν candidate, which is modelled with missing energy from the event (MET see Section 7.7), and applying a constraint based on the W mass. Jet candidates are then selected such that m_{total} is within a top mass window where m_{total} is the sum of the jet, lepton and ν masses. In the case that no leptonic top jet candidate is found it is not set and then all jets are re-introduced to the pool of jets available for the hadronic top. In the case that no hadronic top is found to satisfy the mass criteria then the three hardest available jets are selected.

8.2.2 DNN VR studies

In this sub-section Data/MC comparisons are shown for DNNs input variables (17 in total) and the output of $P(t)$. The distributions are studied in a validation region that is completely orthogonal to the boosted signal region. The aim of this study was to look at the data/MC agreement of the DNN, not to probe $t\bar{t} + b$ modelling, and to check that discrepancies are covered by the fit model. The selection uses the standard ATLAS jet/lepton definitioans and is defined as:

- $= 1$ lepton with $p_T > 27$ GeV and $|\eta| < 2.5$,
- ≥ 4 jets with $p_T > 25$ GeV and $|\eta| < 2.5$,
- ≥ 1 reclustered jet with $p_T > 200$ GeV and $|\eta| < 2.0$,
- ≥ 1 boosted top,
- $= 0$ boosted Higgs,

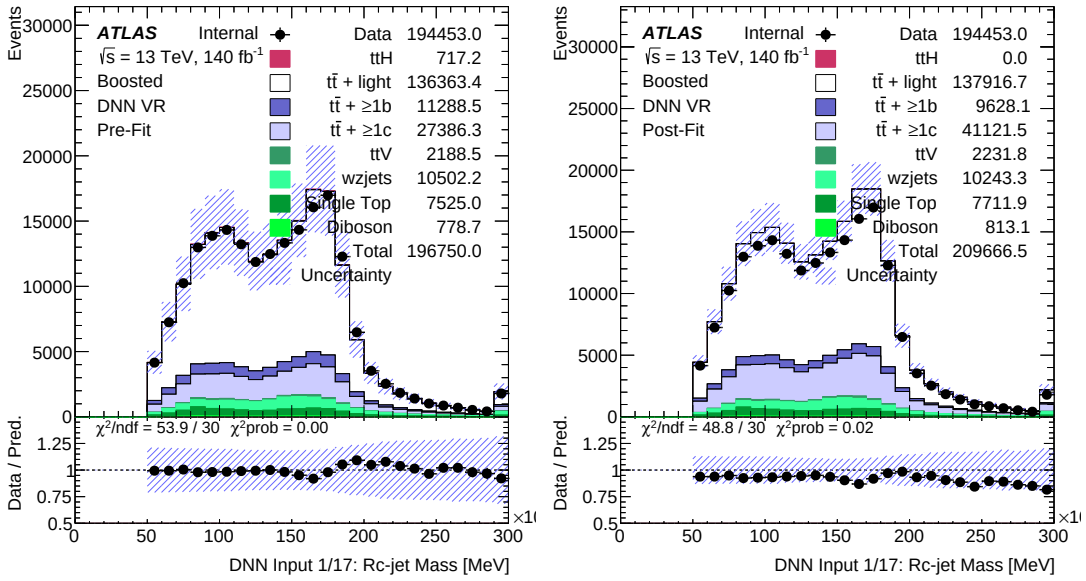
Here a tag by the boosted DNN means that the maximum class score is selected i.e $P(H) > P(t)$, $P(H) > P(QCD)$ for Higgs candidates and $P(t) > P(H)$, $P(t) > P(QCD)$ for top candidates, see previous Subsection 8.2.1 for details on the DNN.

The result of the background only combined fit on data for the total cross-section was applied to this region (Subsection 8.2.3). Some of the smaller uncertainties were not-included: no specific $t\bar{t}+1b$, $t\bar{t}+2b$, Higgs p_T and no renormalisation of $t\bar{t}$+b uncertainties. The nominal $t\bar{t}+b$ model used is a 5 flavour scheme sample with 4 flavour scheme sample being used for the modelling uncertainties.

The validation region heavily selects $t\bar{t}$ events (88.96%) with $t\bar{t}+b = 5.74\%$, $t\bar{t}+c = 13.92\%$ and $t\bar{t}+light = 69.31\%$. It is dominated by $t\bar{t}+light$ which the fit does not provide a strong handle on. Fakes are assumed to be negligible.

It can be seen that for all the DNN inputs the Data/MC agreement looks good and the full list can be seen in Appendix C in Figures C.2-C.6 which show the pre-fit Data/MC comparisons. However, for a few distributions, Mass (Figure 8.14a), d_{23} (Figure 8.16a) and Q_W (Figure 8.18a), some shape effects can be observed and have been highlighted in this section.

For the mass distribution (Figure 8.14a) peaks around the W and top masses can be seen. However, a shape effect can be observed around 170 GeV. In order to investigate this further distributions of the systematic uncertainties were looked at in Figure 8.15 for $t\bar{t}+light$, the dominant MC sample. From this similar shape effects can be observed in the tails of the distribtuions. Jet energy scale uncertainties are expected to be significant because small-r jet energies largely define the mass of the reclustered jet.



(a) (b)

Figure 8.14: Pre-fit (a) and Post-fit (b) DNN input, RC-jet mass.

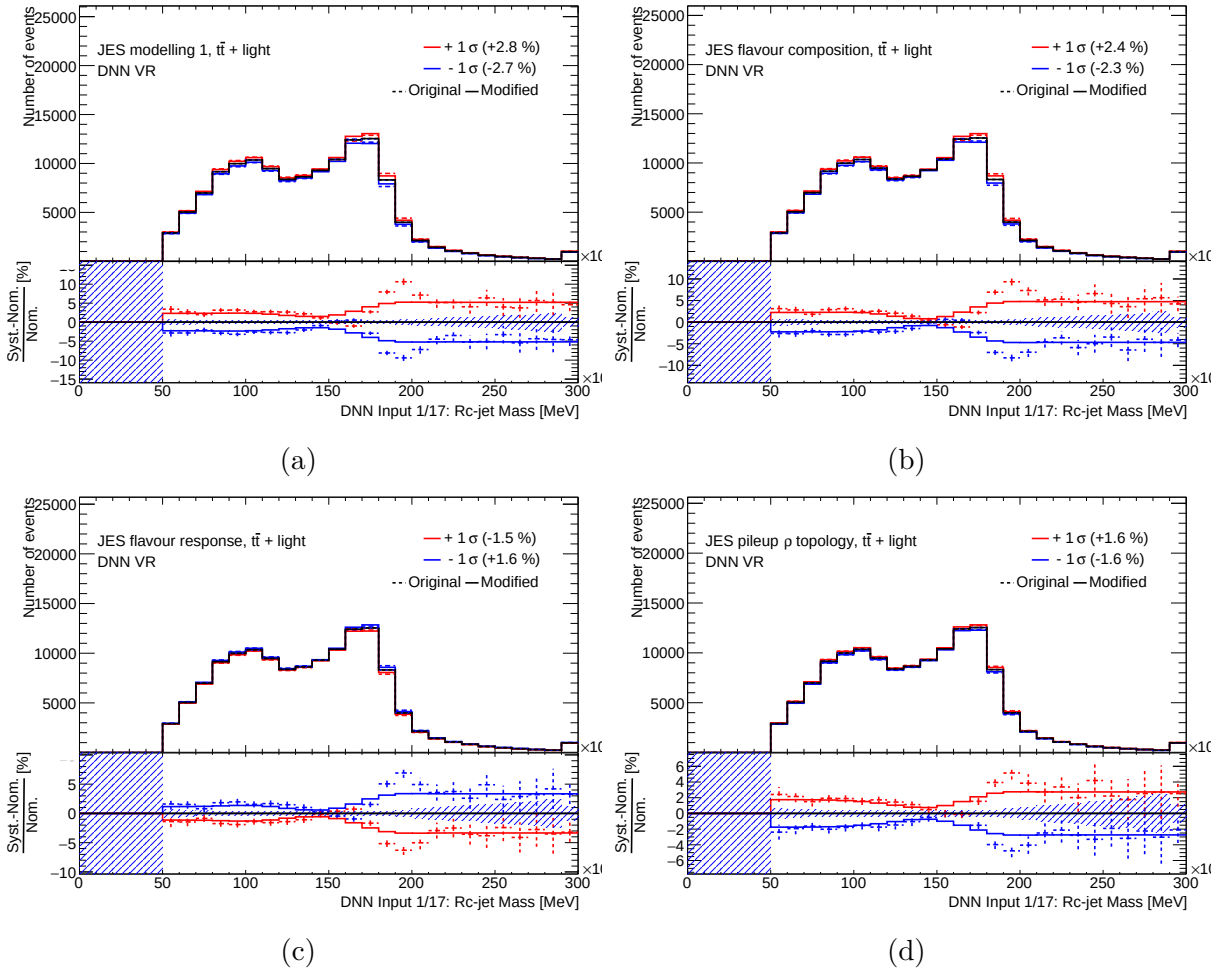
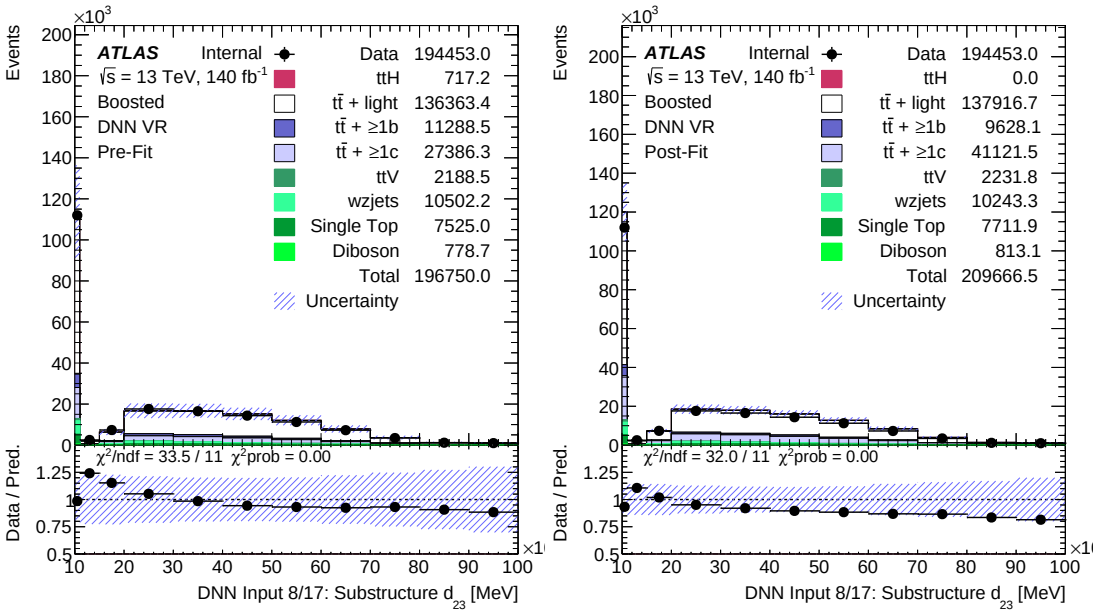


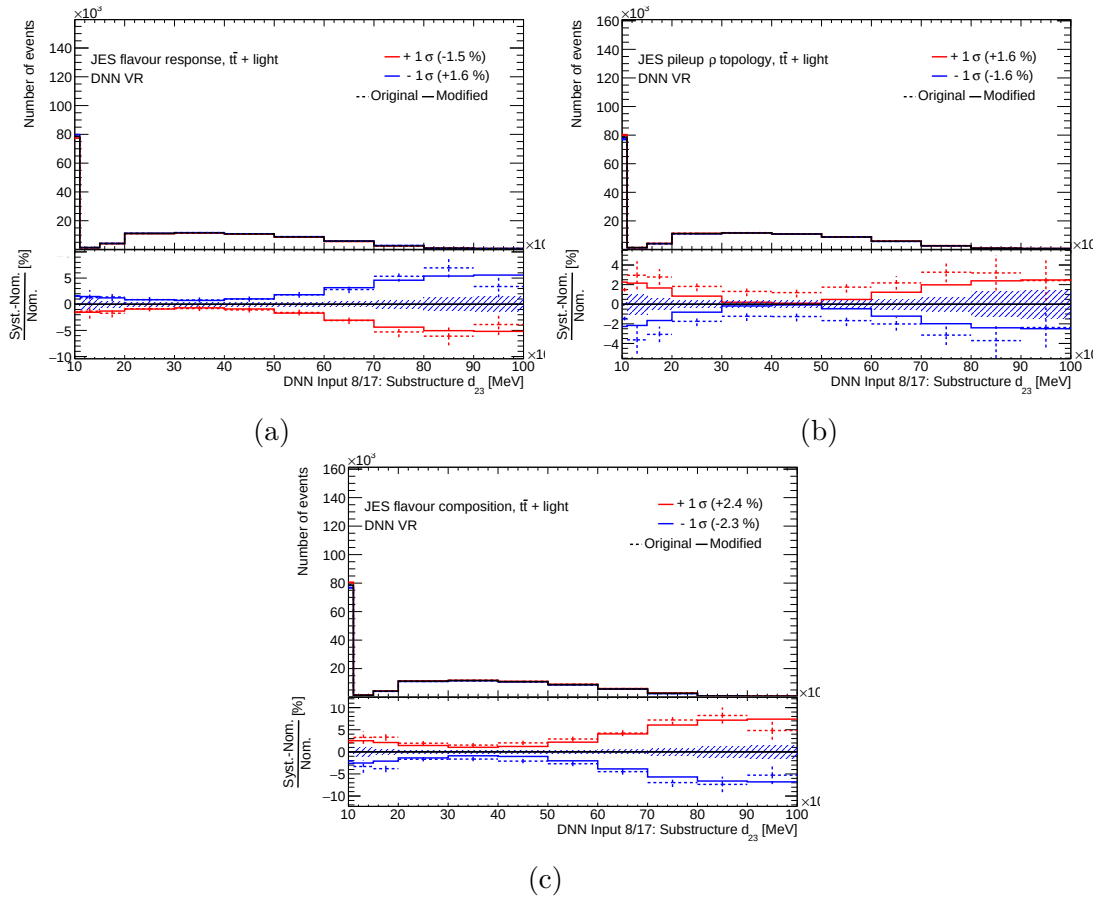
Figure 8.15: Jet energy scale uncertainties for the reclustered jet mass distribution (Figure C.2a): JES modelling 1 (8.15a), JES flavour composition (8.15b), JES flavour response (8.15c) and JES pileup ρ topology (8.15b).

JES uncertainties were also found to be relevant for shape effects observed in the substructure variables d_{23} (Figure 8.16a) and Q_W (Figure 8.18a). For d_{23} the 1st bin (at 0) is filled with reclustered jets that have 2 subjets for which this variable is undefined. The corresponding JES uncertainties (Figure 8.17). For Q_W the corresponding JES uncertainties can be seen in Figures 8.19. In all cases no significant shape effects were found, with them predominantly appearing in the tails of the distributions.



(a) (b)

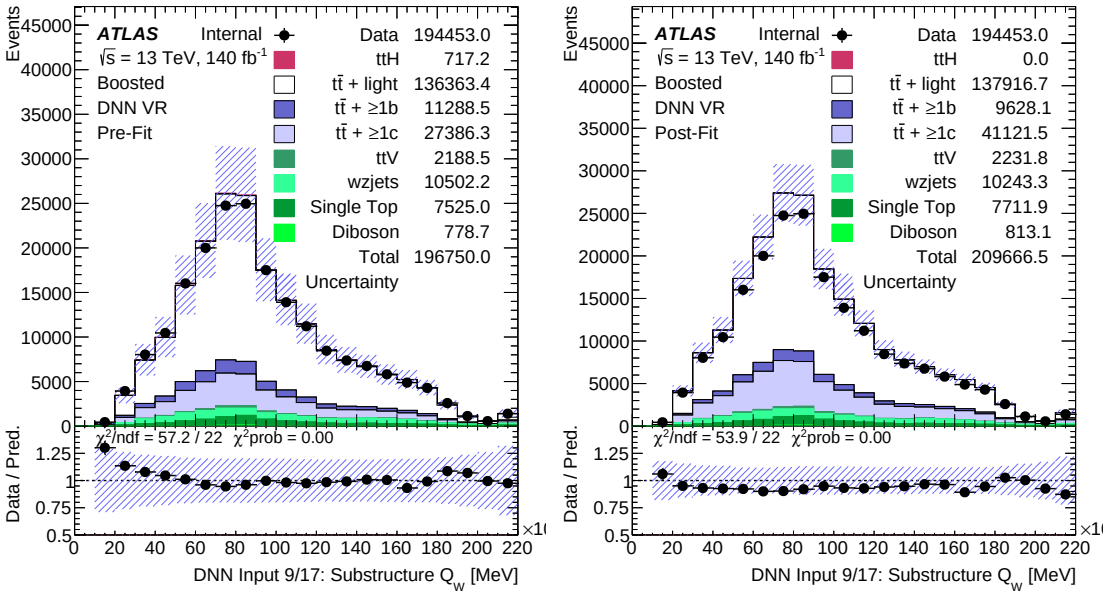
Figure 8.16: Pre-fit (a) and Post-fit (b) DNN input, RC-jet substructure variable d_{23}.



(a) (b)

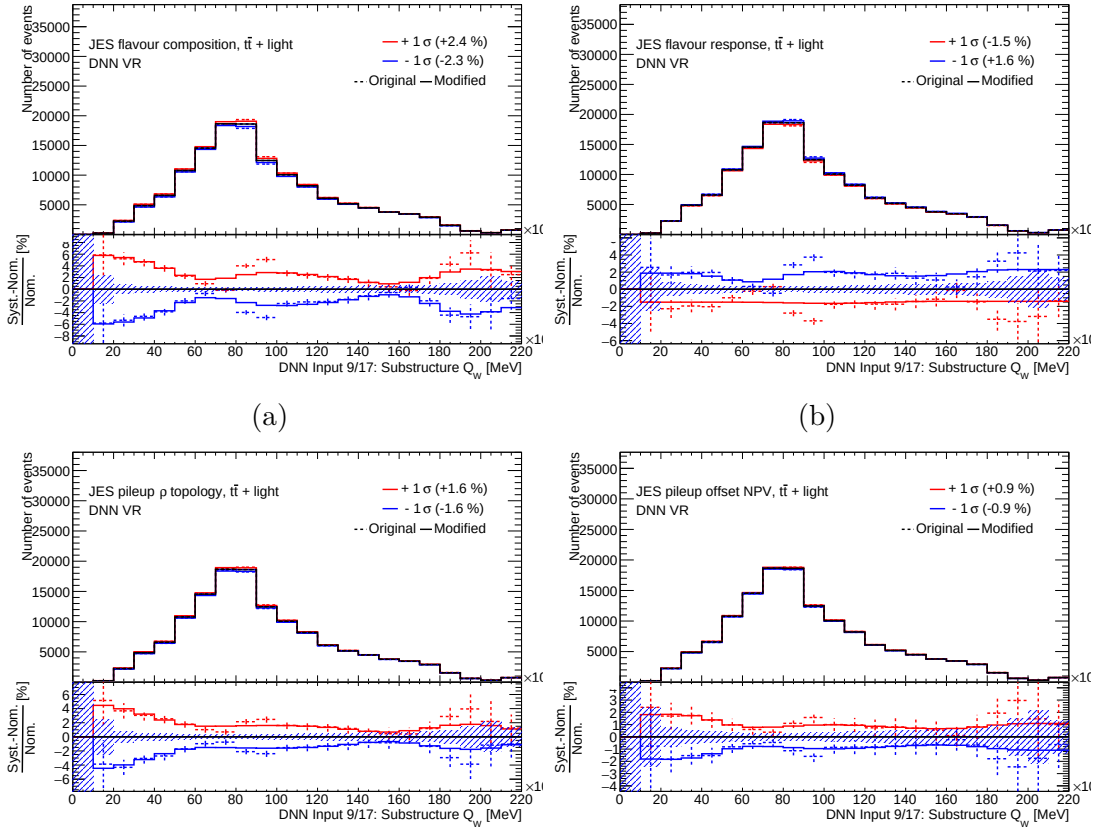
(c)

Figure 8.17: Jet energy scale uncertainties for the substructure variable d_{23} distribution (Figure C.3d): JES flavour response (8.17a), JES JES pileup ρ topology (8.17b) and JES flavour composition (8.17c).



(a) (b)

Figure 8.18: Pre-fit (a) and Post-fit (b) DNN input, RC-jet substructure variable Q_w.



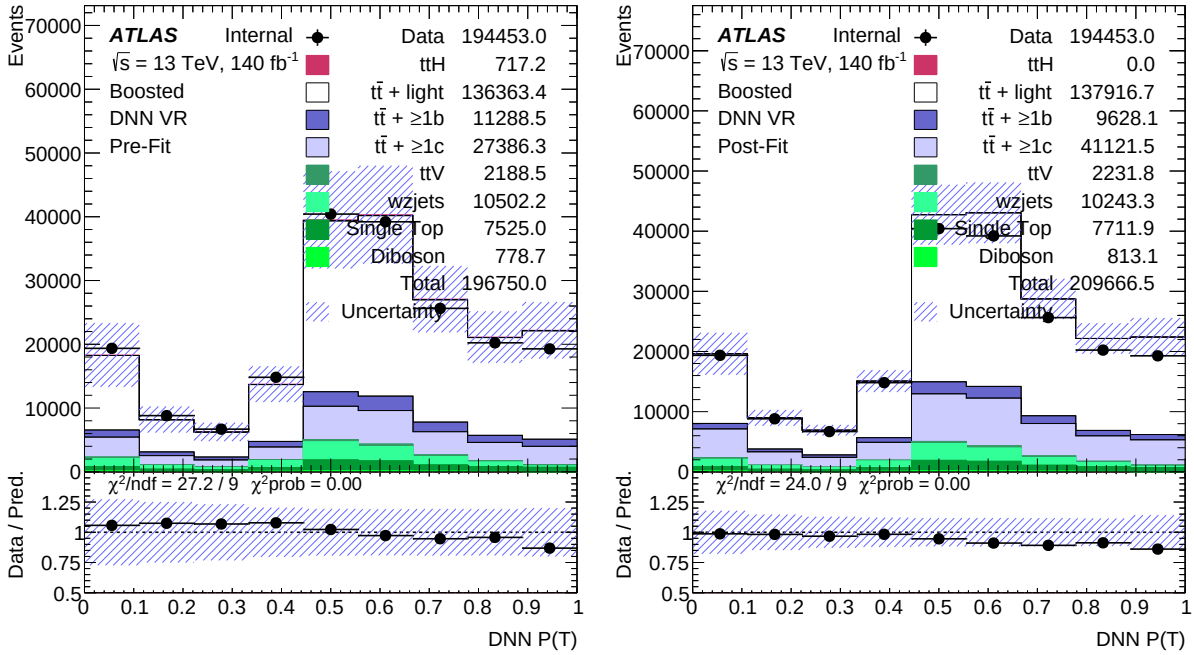
(a) (b)

(c) (d)

Figure 8.19: Jet energy scale uncertainties for the substructure variable Q_W distribution (Figure C.4a):JES flavour composition (8.19a), JES flavour response (8.19b), JES pileup ρ topology (8.19c) and JES pileup offset NPV (8.19d).

For all the DNN inputs the pre- and post-fit plots can be seen in Appendix C.3 in Figures C.7-C.11. Overall the Data/MC agreement for all distributions is within the post-fit uncertainty bands.

The pre- and post-fit plots for the DNN output $P(t)$ can be seen in Figure 8.20. Some QCD tagged reclustered jets can be seen in the region taking $P(t)$ values between 0 and 0.333. A slant in the Data/MC agreement can be seen from $0.4 \rightarrow 1.0$ but is within the uncertainties. The largest contribution is from $t\bar{t}+light$ with $t\bar{t}+c$ being scaled up $\sim 50\%$ by the fit. Again the Data/MC agreement is within these post-fit uncertainty bands.



(a) (b)

Figure 8.20: Pre-fit (a) and Post-fit (b) DNN output RC-jet $P(t)$.

The modelling of $t\bar{t}+light$ (Figure 8.21, 8.22) was looked at as it's the largest MC sample, we can see that these modelling uncertainties are some of the largest. The modelling of $t\bar{t}+c$ (Figure 8.23, 8.24)) was also included as this sample is both the 2nd largest contribution and is scaled up by the fit. Overall it can be seen that these uncertainties are not significant and do not show anything out of the ordinary.

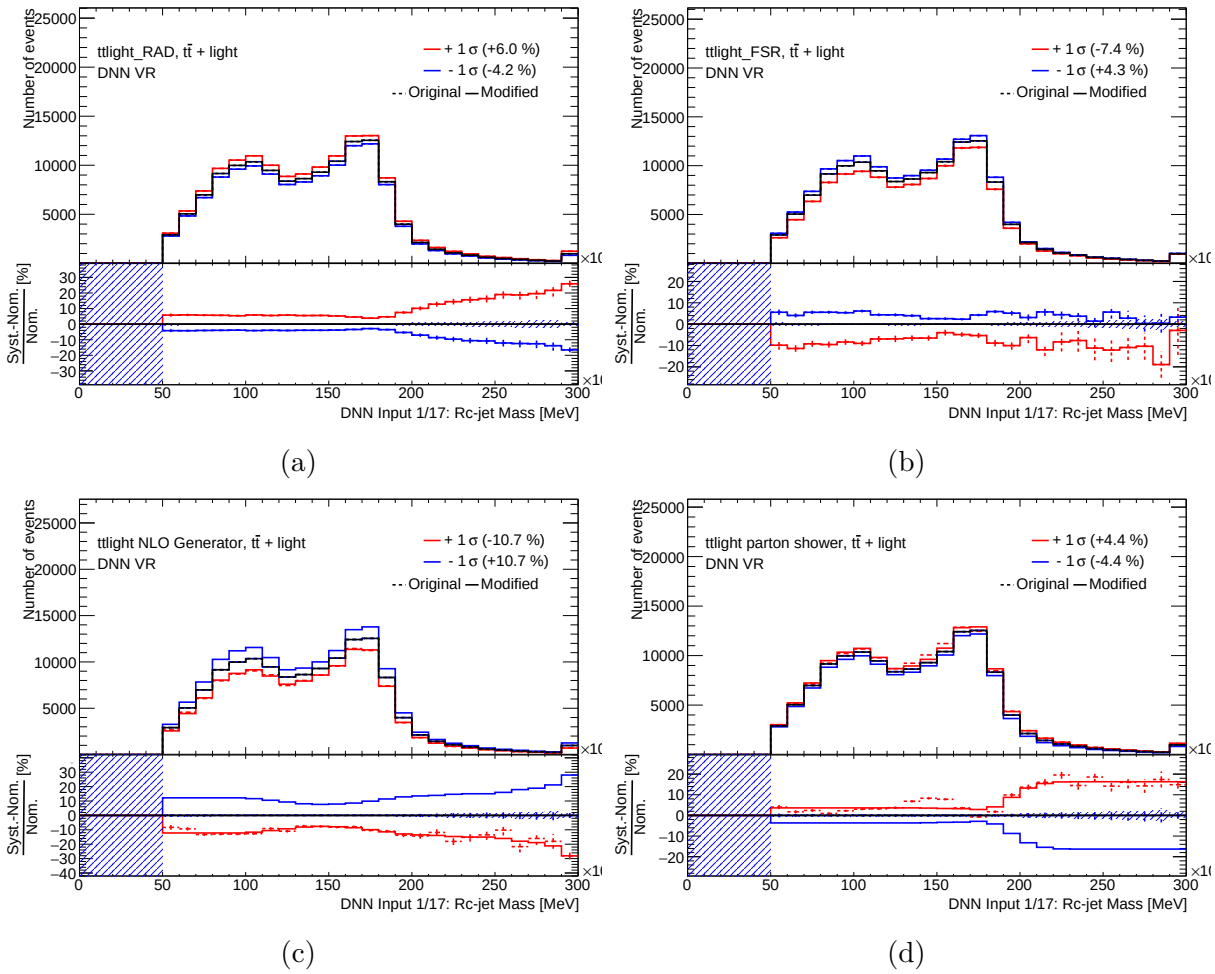


Figure 8.21: $t\bar{t} + light$ systematic modelling uncertainties for the reclustered jet mass distribution: initial state radiation (8.21a), final state radiation (8.21b), NLO generator (8.21c) and parton shower (8.21d).

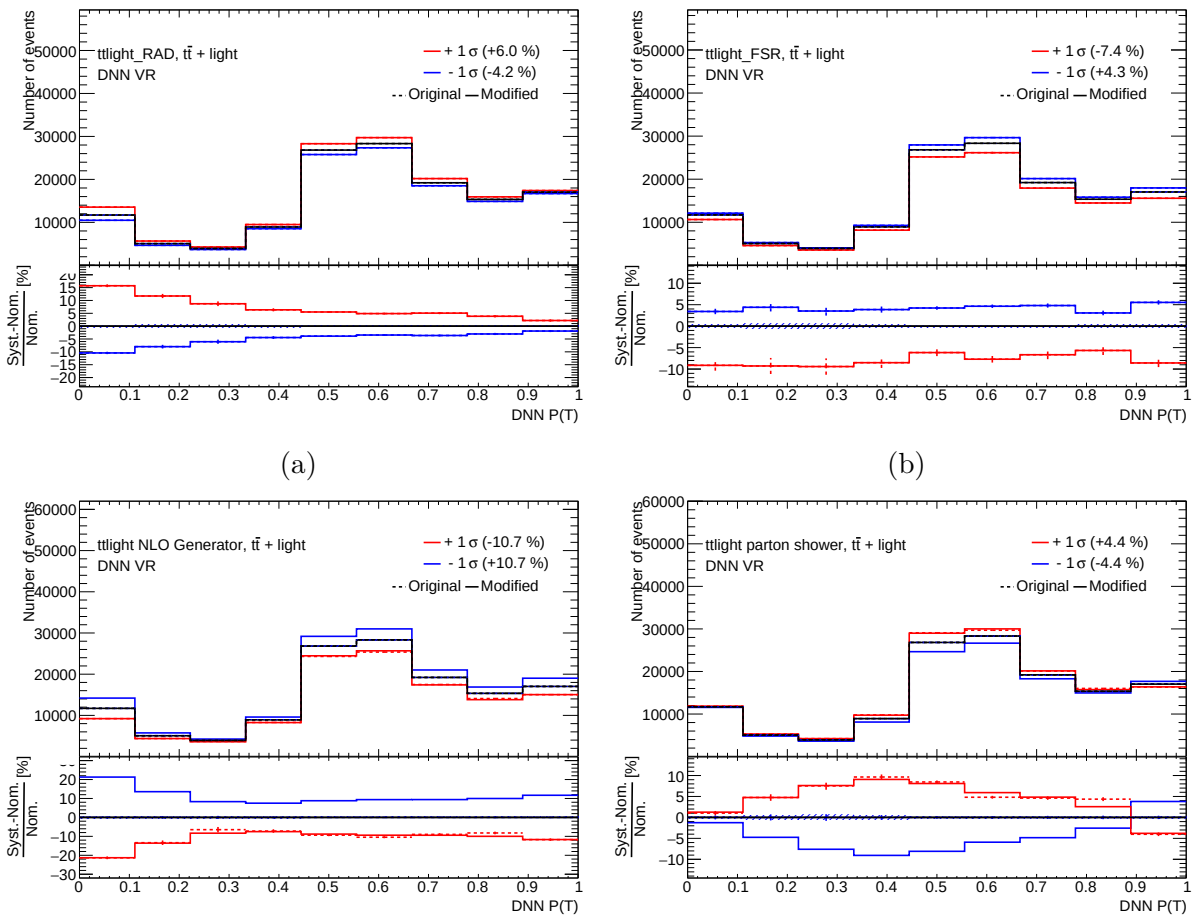


Figure 8.22: $t\bar{t}+light$ systematic modelling uncertainties for the DNN output $P(t)$ distribution: initial state radiation (8.22a), final state radiation (8.22b), NLO generator (8.22c) and parton shower (8.22d).

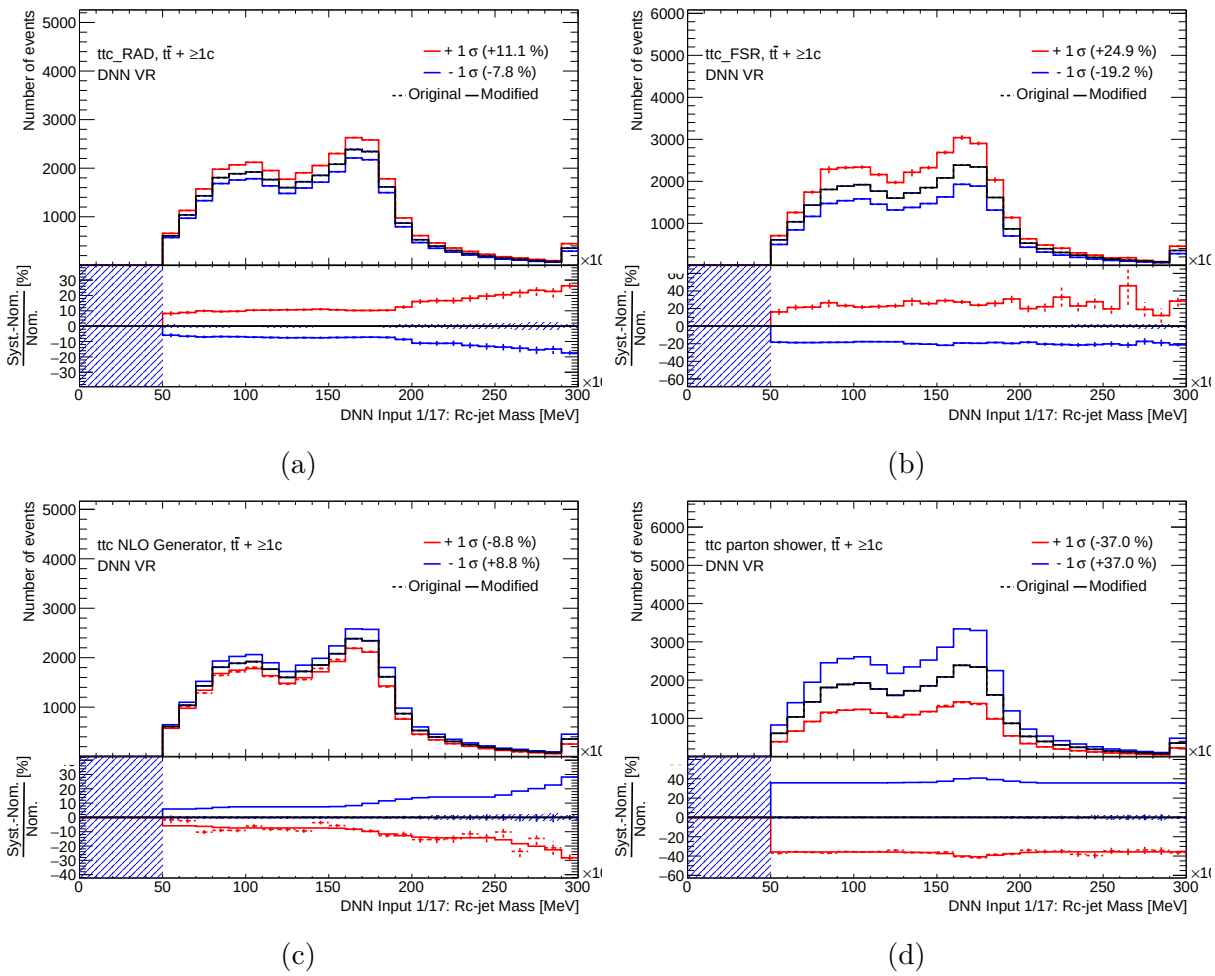


Figure 8.23: $t\bar{t} + c$ systematic modelling uncertainties for the reclustered jet mass distribution: initial state radiation (8.23a), final state radiation (8.23b), NLO generator (8.23c) and parton shower (8.23d).

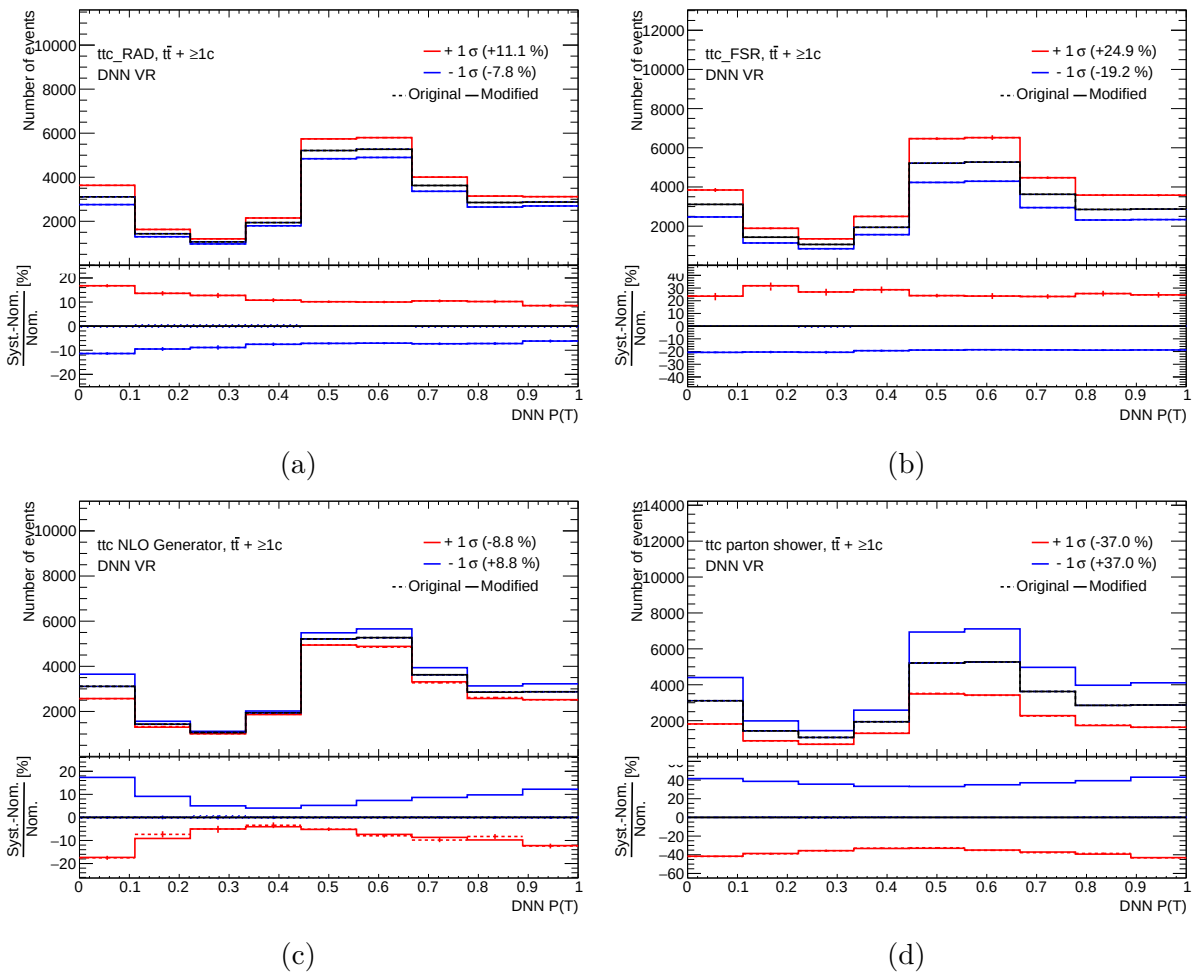


Figure 8.24: $t\bar{t}$+c systematic modelling uncertainties for the DNN output $P(t)$ distribution: initial state radiation (8.24a), final state radiation (8.24b), NLO generator (8.24c) and parton shower (8.24d).

Overall the DNN input variables and output are well modelled in this validation region. However, the χ^2 values are not always good e.g. in the invariant mass of b-tagged subjet distributions (Figures C.4b, C.9b) and both have a χ^2prob $= 0$. In this case it is not much of an issue as there is a low b-jet multiplicity in this region. Some shape effects can be noticed for invariant mass quantities with them also showing up in the uncertainties.

8.2.3 Results

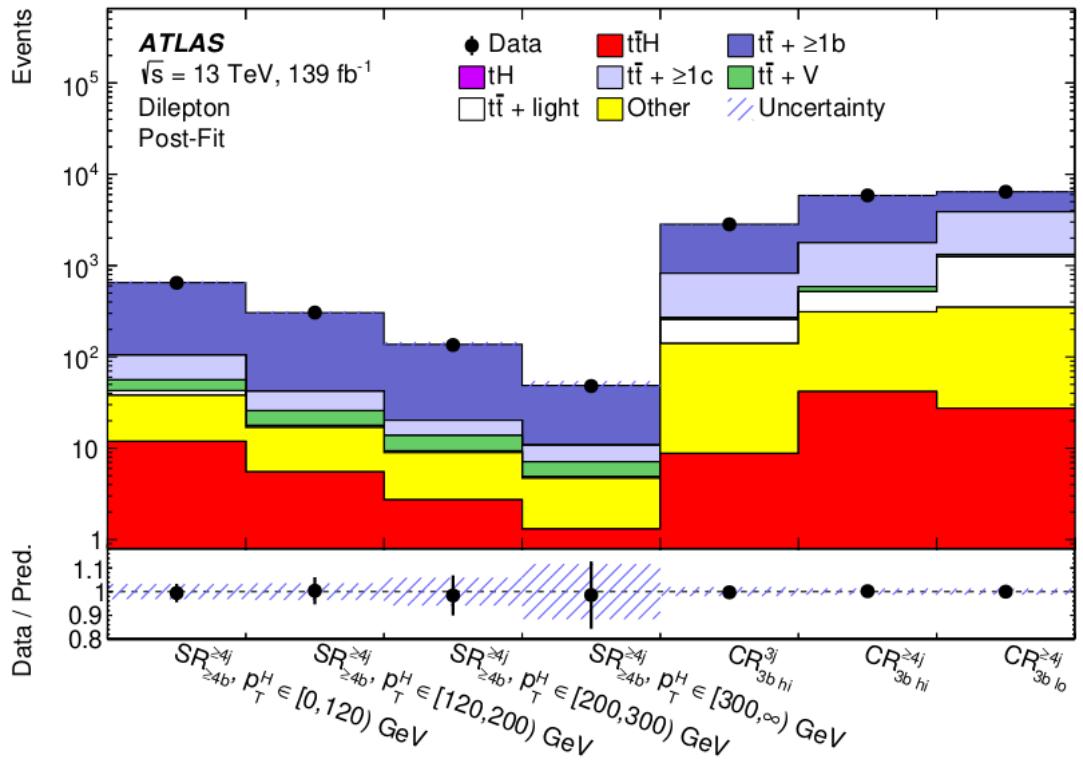
The $t\bar{t}H$ event classification BDT distributions for each SR, along with the event yields and ΔR_{bb}^{avg} for the CRs (as defined in Table 8.5) are combined in a profile likelihood fit using the TRex Fitter package [106] which uses implementations from the RooStat framework [107, 108]. The pre-fit and post-fit event yields for the boosted region can be seen in Table 8.6 with a breakdown for each nominal signal / background MC sample. The fit dramatically scales down the $t\bar{t}H$ event yield while scaling up $t\bar{t}+ \geq 1c$. This is

observed for both the boosted region (see Table 8.6) as well as for the resolved regions (see paper [1]). The boosted region in particular is focused on because it is more sensitive to new physics.

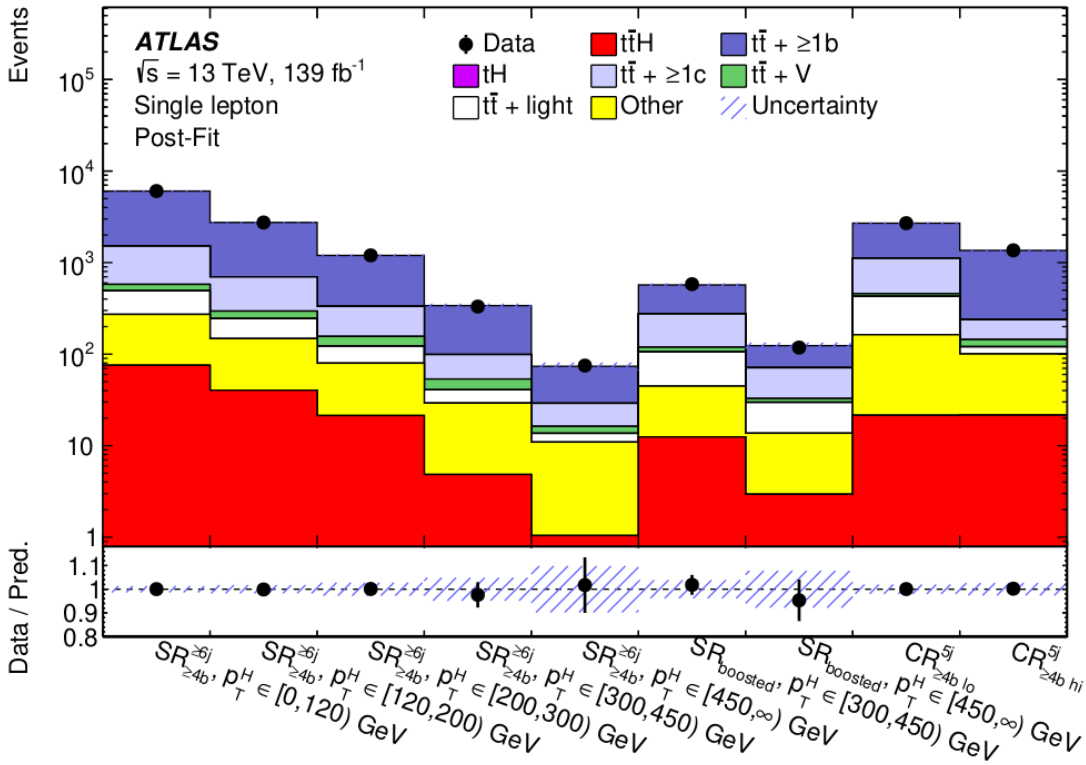
	$SR_{boosted}$, $p_T^H \in [300,450)$ GeV		$SR_{boosted}$, $p_T^H \in [450,\infty)$ GeV	
	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t}H$	35.1 ± 4.1	12 ± 12	8.5 ± 1.1	3.0 ± 2.9
tH	1.31 ± 0.31	1.31 ± 0.29	0.42 ± 0.10	0.42 ± 0.10
$t\bar{t}+\geq 1b$	246 ± 49	295 ± 23	55 ± 24	52.5 ± 9.3
$t\bar{t}+\geq 1c$	84 ± 90	156 ± 35	21 ± 23	38 ± 10
$t\bar{t}+W$	1.86 ± 0.39	1.88 ± 0.35	0.55 ± 0.18	0.56 ± 0.17
$t\bar{t}+Z$	10.7 ± 2.1	10.9 ± 2.1	2.21 ± 0.60	2.32 ± 0.59
$t\bar{t}$ + light	56 ± 26	61 ± 25	17 ± 10	16.2 ± 7.3
Single top tW	13.1 ± 8.0	13.7 ± 8.3	6.1 ± 5.8	5.3 ± 4.6
$t\bar{t}t\bar{t}$	1.76 ± 0.89	1.74 ± 0.88	0.42 ± 0.22	0.41 ± 0.21
Other top sources	4.3 ± 3.2	4.4 ± 3.1	0.80 ± 0.78	0.82 ± 0.77
V & VV + jets	12.4 ± 5.7	12.4 ± 5.4	4.3 ± 2.3	4.2 ± 2.0
Total	470 ± 120	571 ± 22	117 ± 38	123.8 ± 9.7
Data		581		118

Table 8.6: Pre-fit and post-fit event yields for boosted single-lepton SRs. The post fit yields are after the inclusive fit in all channels with all uncertainties being included in both pre-fit and post-fit uncertainties apart from the $k(t\bar{t}+\geq 1b)$ normalisation factor which is only defined for post-fit [1].

Post-fit values for each SR + CR (for both dilepton and single-lepton) are summarised in Figure 8.25, it can be observed that the uncertainties increase as a function of Higgs p_T. The effectiveness of the single-lepton boosted region can be seen when looking at the two STXS bins Higgs $p_T \in [300, 450) \cup [450, \infty)$ for both resolved and boosted single-lepton regions. The boosted techniques utilised capture a good portion of the $t\bar{t}H$ events from the resolved region with smaller uncertainties for equivalent bins.



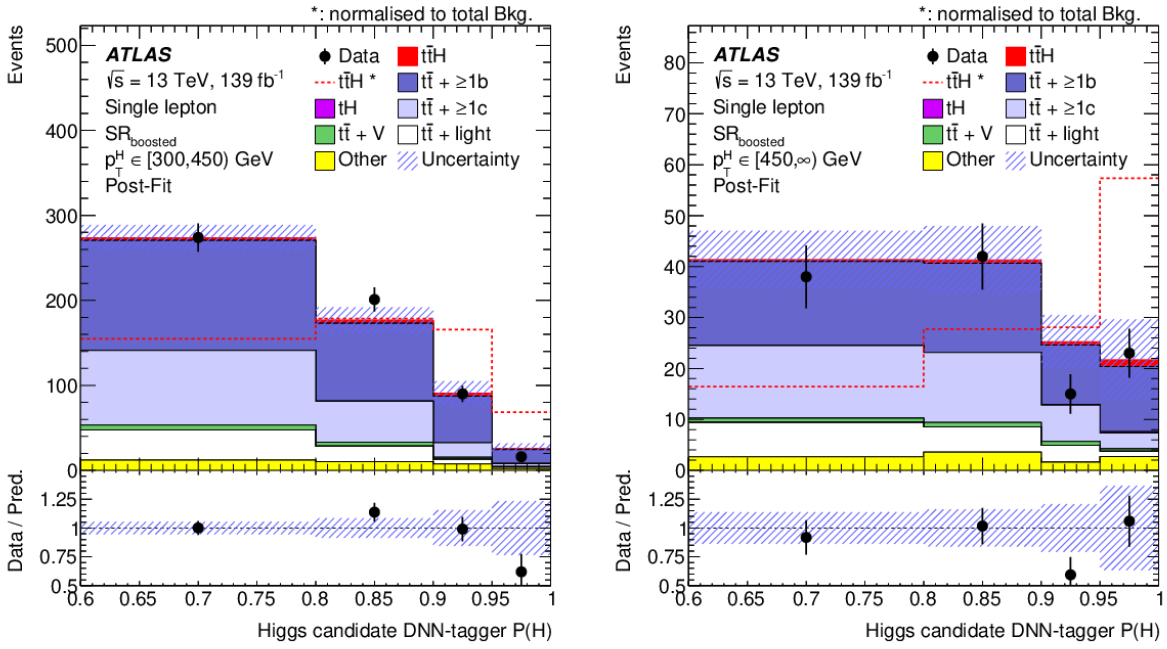
(a)



(b)

Figure 8.25: Predicted and observed event yield comparison between all regions defined in Table 8.5 for dilepton (a) and single-lepton (b) [1].

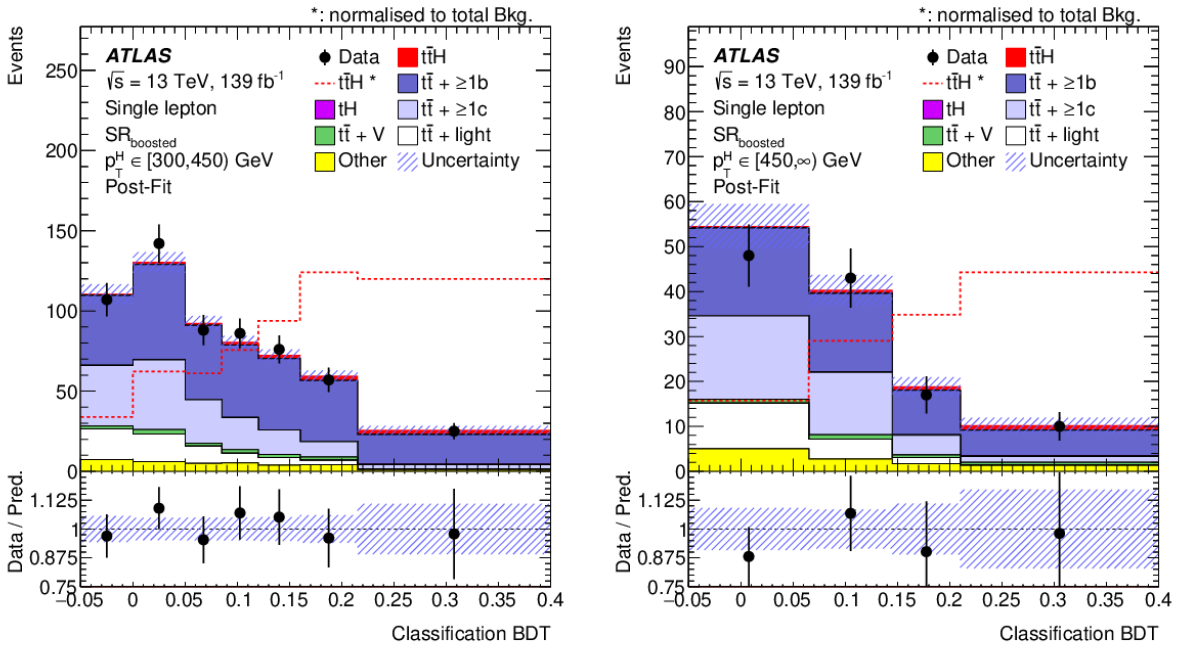
The boosted-single lepton post-fit DNN can be seen in Figure 8.26 for both STXS bins. We can see from the dashed red line representing the signal scaled to the total background, for shape comparisons, that for the STXS $p_t \in [300, 450)$ bin that the signal events heavily drop off in the final DNN $P(H)$ bin. This is mirrored by a drop in the Data/MC agreement for that bin. In the STXS $p_t \in [450, \infty)$ bin the opposite behaviour can be observed with the signal events being maximally selected for the same final DNN $P(H)$ and have good modelling. This behaviour could be due to the DNN working better at higher p_T where the decay products are more collimated and so better encapsulated by the RC-jet. These DNN $P(H)$ distributions then enter into the boosted classification BDT.



(a) (b)

Figure 8.26: Data/MC agreement for boosted sing-lepton Higgs reconstruction DNN for the two STXS bins Higgs $p_T \in [300, 450) \cup [450, \infty)$ [1].

The boosted-single lepton post-fit BDT can be seen in Figure 8.27 for both STXS bins. We can see from the dashed red line representing the signal scaled to total background, for shape comparisons, that it does indeed preferentially select signal events. In the final bin the only significant background contamination comes from the signal-like $t\bar{t} + \geq 1b$ background which is notoriously hard to separate. The overall Data/MC agreement for the classification BDT is good post fit.



(a) (b)

Figure 8.27: Data/MC agreement for boosted sing-lepton $t\bar{t}H$ event classification BDT for the two STXS bins Higgs $p_T \in [300, 450) \cup [450, \infty)$ [1].

From Figure 8.28 we can see that the boosted Higgs candidates invariant mass is well modelled and within the uncertainty bands. In both cases (Figure 8.14 for top candidates mass) the fit corrected the Higgs / top candidates mass within uncertainties and shows good modelling overall.

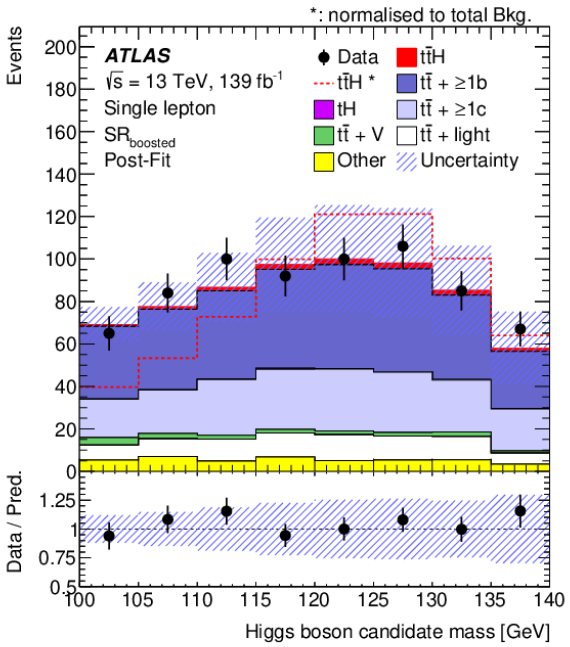


Figure 8.28: Boosted single-lepton Higgs candidate invariant mass distribution [1].

The results of the inclusive fit and the fit for the dilepton, resolved and boosted single-lepton regions can be seen in Figure 8.29. The fit is repeated for each signal channel: resolved single-lepton, boosted single-lepton and dilepton. Thus obtaining a signal strength μ for each channel where $\mu = \sigma^{t\bar{t}H}_{obs}/\sigma^{t\bar{t}H}_{SM}$. The best-fit value for $\mu = 0.35$ with ± 0.20 statistical uncertainty and $^{+0.30}_{-0.28}$ systematic uncertainty giving $\mu = 0.35^{+0.36}_{-0.34}$. This corresponds to $\sigma = 1.0(2.7)$ observed(expected) significance compared to the background-only hypothesis. The global goodness of fit, which includes all the input variables to the classification BDTs, is 92% indicating good post-fit modelling. The SM compatibility of 8.5% prediction is made using the Asimov fit where μ is fixed to 1. Overall the result is compatible with the [23] $36.1 fb^{-1}$ result a subset of the full ATLAS Run-2 data set, with the impact of the systematics being reduced by a factor of two.

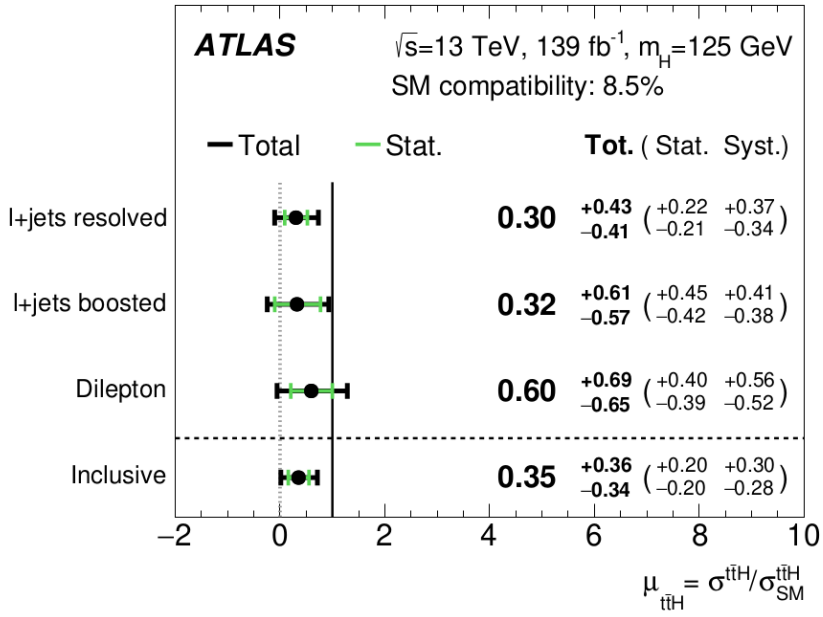


Figure 8.29: Fitted values of $t\bar{t}H$ signal strength for each region given both separately and combined for the inclusive measurement [1].

The largest pulls on systematic observables by the fit can be seen in Figure 8.30 which is dominated by $t\bar{t}+ \geq 1b$ systematics. The largest pull $\approx 1.2\sigma$ is on the $t\bar{t}+ \geq 1b$ ISR uncertainty which is mostly driven by the renormalisation scale. This indicates that the data favours a softer renormalisation scale in the ME calculation. It was shown that this pull corrects the jet multiplicity (by adjusting the additional radiation) in the SRs, which is shown for the boosted region in Figure 8.31. It was also shown that when decorrelating this parameter between single-lepton and dilepton they both re-produced the same pull with μ effectively unchanged.

The reconstructed shape uncertainty p_T^{bb} for the $t\bar{t} + \geq 1b$ background sample also exhibits a large uncertainty, as previously mentioned and expected from the miss-modelling present in the Higgs candidates p_T present in the pre-fit Figure 8.13c. This uncertainty is defined such that a $+1\sigma$ variation corresponds to correcting the reconstructed Higgs p_T shape by adjusting the background model as can be seen in Figure 8.32 for the boosted Higgs candidate. The effect of this was assessed by replacing this with decorrelated free-floating $t\bar{t}+ \geq 1b$ normalisation factors across the STXS bins (which have different sensitivities to different decay modes) and regions with no bias being observed.

The normalisation factor for $t\bar{t}+ \geq 1b$ background is left free floating in the fit and is found to be $k(t\bar{t}+ \geq 1b) = 1.28 \pm 0.08$ as can be seen in Figure 8.30. The $t\bar{t}+ \geq 1b$:NLO match. dilep $p_T^H \in [0, 120)$ GeV uncertainty is the final systematic with a pull close to 1σ and is ranked 6th in its impact. To investigate this a fit was performed with an alternative $t\bar{t}+b\bar{b}$ MC sample using the Sherpa 2.2.1 generator with no significant bias being observed. Due to the number of nuisance parameters used in the fit there is a good chance that any

of the $t\bar{t}+ \geq 1b$ NLO matching uncertainties can be pulled to an extreme value when looking from a purely statistically standpoint, this is calculated to be 17%.

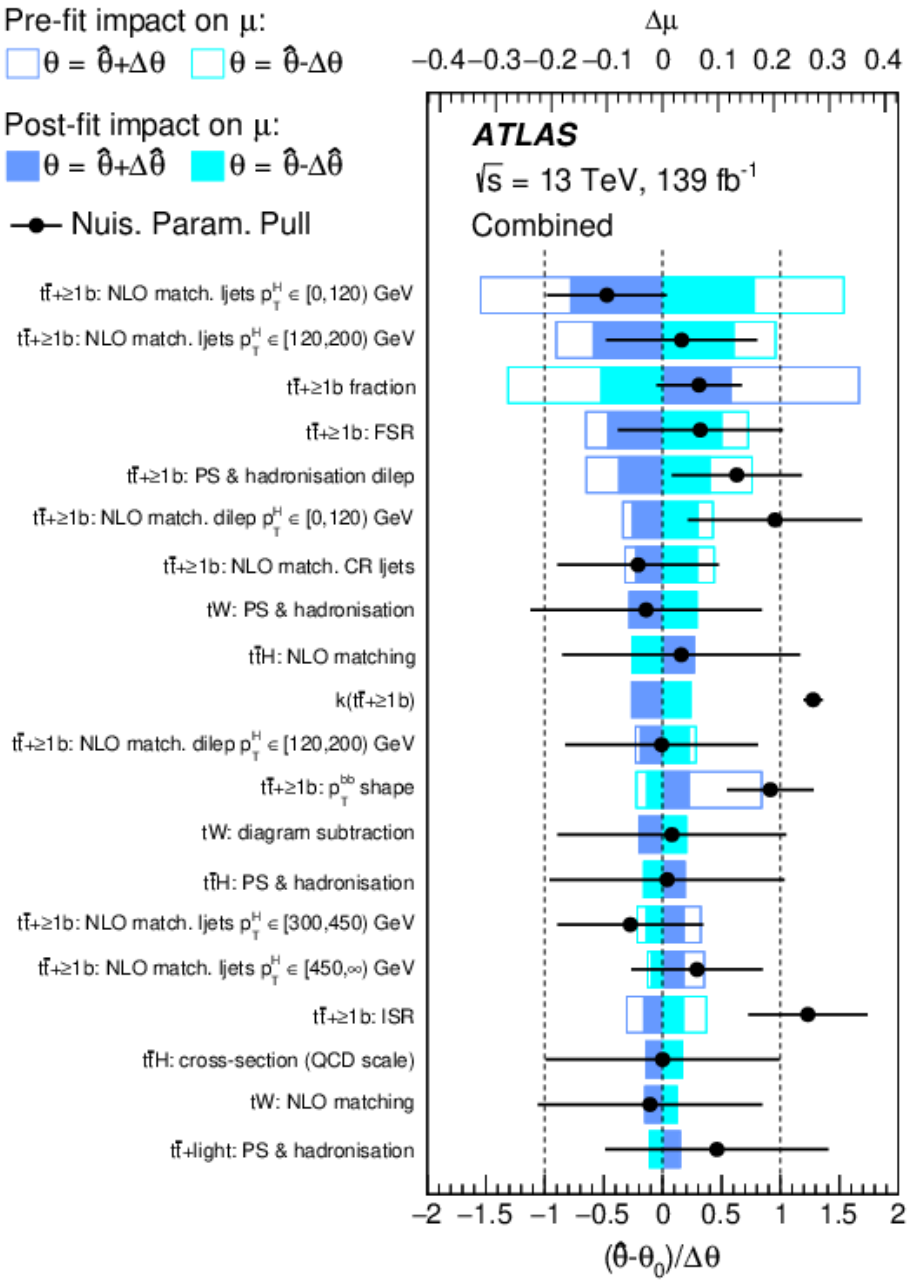
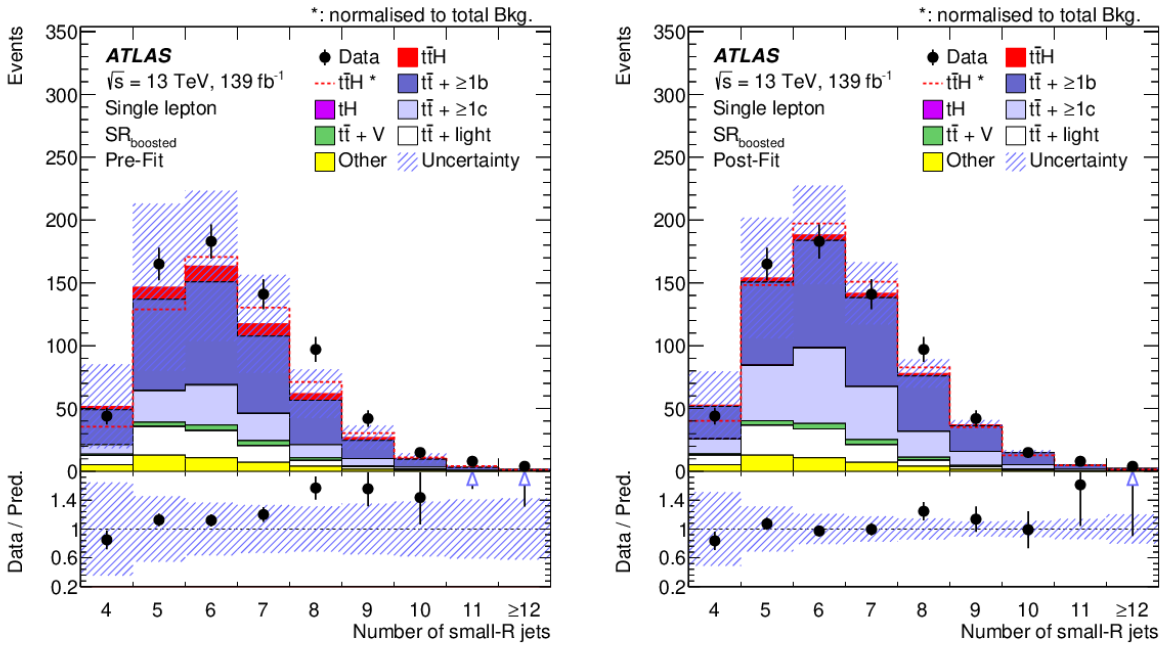


Figure 8.30: Ranking plot of the 20 nuisance parameters with the largest post-fit impact on μ in the fit. Statistical uncertainty nuisance parameters are not included. The blue rectangles define the impact on μ with the blue outlines referring to pre-fit impact and the solid blue referring to the post-fit impact. Impacts are calculated by fixing and then fitting each nuisance parameter individually to its best fit value shifted by the pre(post)-fit uncertainties and comparing to the nominal best-fit value of μ [1].



(a) (b)

Figure 8.31: Pre-fit (a) and post-fit (b) jet multiplicity distributions for the boosted SR, highlighting the $t\bar{t}+ \geq 1b$ ISR correction [1].

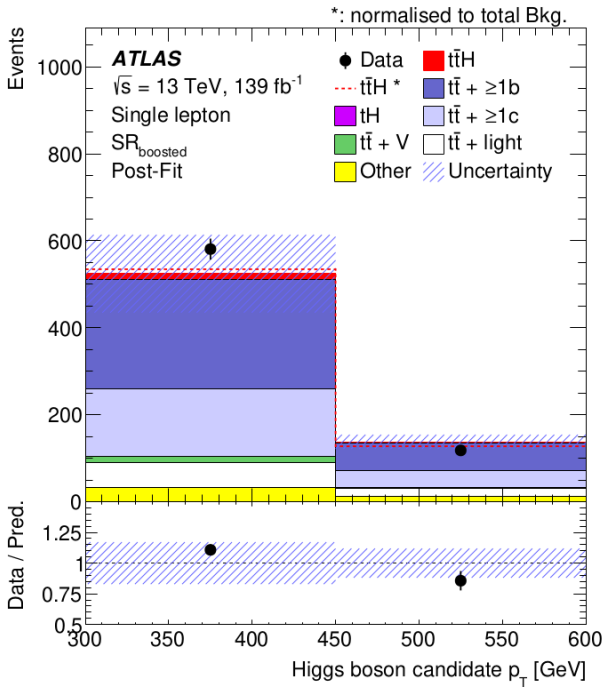


Figure 8.32: Post-fit p_T distribution for the boosted Higgs candidate [1].

The combined fit is repeated for each combined STXS bin in order to obtain a μ signal strength for each. The results of this can be seen in Figure 8.33 with the same fit parameters entering into the fit as before with the inclusive fit but in the 11 STXS bins defined for the regions. The free floating normalisation factor is found to be $k(t\bar{t}+ \geq 1b) = 1.28 \pm 0.08$ in agreement with the inclusive fit. The global goodness of fit is 88% indicating good post-fit modelling with a SM compatibility of 45%, which in similar fashion to before was obtained by fixing the μ in each final STXS bin to 1. Overall the fit has similar behaviour to the inclusive fit, broadly matching the pulls and constraints on the nuisance parameters. The first combined STXS bin ($p_T \in [0, 120)$ GeV) is dominated by $t\bar{t}+ \geq 1b$ modelling systematics with statistical uncertainties dominating in the rest.

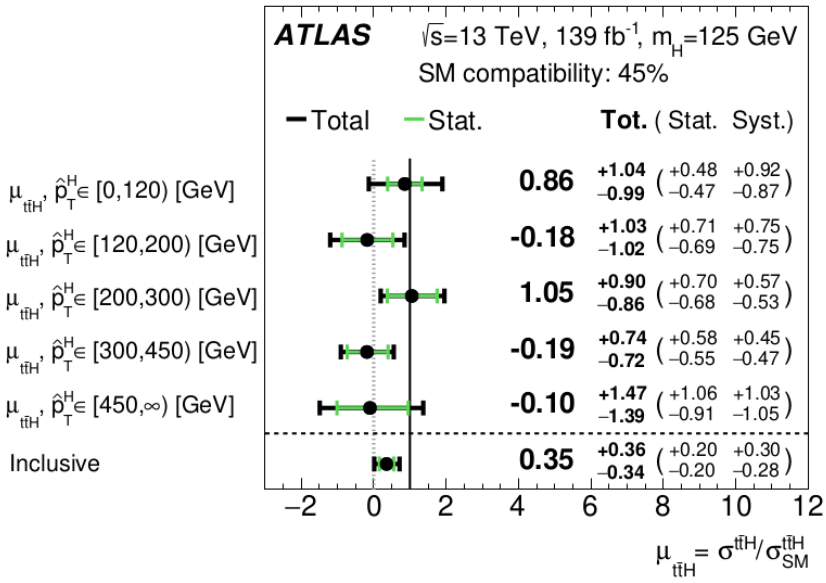


Figure 8.33: Fitted values of $t\bar{t}H$ signal strength for each combined STXS bin and the inclusive signal strength [1].

The cross-section upper limits are also derived for each combined STXS bins and can be seen in Figure 8.34 where the 95% confidence levels are shown. For this the likelihood function used for measuring the signal strengths is modified. The PDF and signal scale uncertainties are omitted because they have no impact on the cross-section. Scale effects are still present through the inclusion of ISR uncertainties but have no impact on the overall cross-section. The inclusive cross-section of 507 fb^{-1} is used to calculate the limits with the relevant fraction being used for each combined STXS p_T bin to establish said bins cross-section prediction.

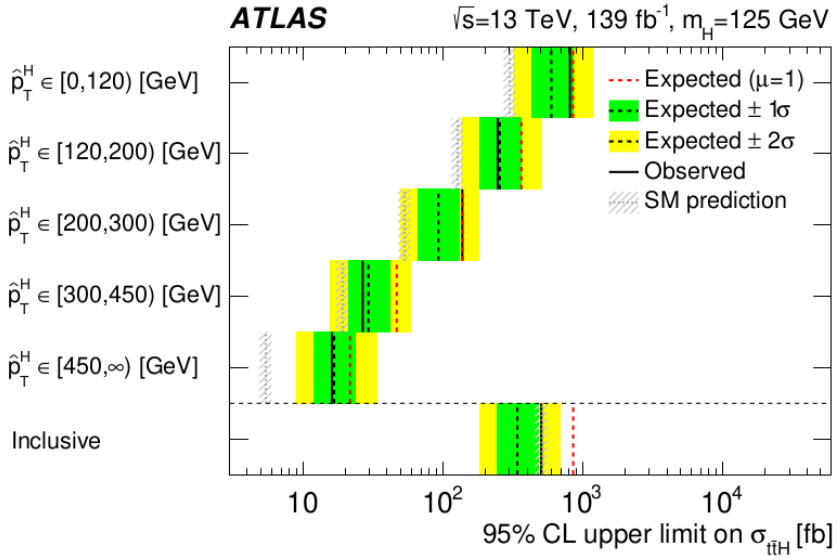


Figure 8.34: 95% CL simplified template cross-section upper limits for each of the combined STXS p_T bins as well as for the inclusive limit [1].

Overall systematic uncertainties are slightly more dominant for the inclusive measurement with $t\bar{t}$ modelling systematics dominating the ranking plot (see Figure 8.30). The boosted region suffers slightly more from statistical uncertainties (see Figure 8.29). The Lepton + Jets Legacy Analysis reuses the ATLAS Run-2 data set, so statistical uncertainties are unexpected to change. However, it is expected to benefit from improved $t\bar{t}$ modelling. The boosted region is also expected to benefit from an improved b-tagging algorithm with better performance at high p_T. Statistical uncertainties should also improve once Run-3 data becomes available with the full combination expected to triple the data available for analyses.

When comparing to the CMS $t\bar{t}H \rightarrow b\bar{b}$ results using: $35.9fb^{-1}$ [24], $41.5fb^{-1}$ and $77.4fb^{-1}$ [25] of data. This corresponds to the 2016, 2017 and 2016-17 combination which are subsets of the CMS full Run-2 data set of $137.24fb^{-1}$. They achieved significances $\sigma = 1.6(2.2)$, $\sigma = 3.7(2.6)$ and $\sigma = 3.9(3.5)$ observed(expected) and measured $\mu = 0.72^{+0.45}_{-0.45}$, $\mu = 1.49^{+0.44}_{-0.40}$ and $\mu = 1.15^{+0.32}_{-0.29}$ respectively. The previous ATLAS result [23] uses a $36.1fb^{-1}$ subset of the full Run-2 dataset and achieves a significance of $\sigma = 1.4(1.6)$ and measured $\mu = 0.84^{+0.64}_{-0.61}$. From this we can see that CMS had and up fluctuation in their data while ATLAS had a down fluctuation.

8.3 Lepton + Jets Legacy Analysis

The ATLAS collaboration continuously develops its analysis techniques including the algorithms and objects used in them. Because the previous result was ultimately dominated by $t\bar{t}$ modelling systematic uncertainties it was decided that a re-analysis of the existing

Run-2 ATLAS data set, referred to as the $t\bar{t}H$ lepton + jets "Legacy" analysis, would be beneficial as improvements in both $t\bar{t}$ modelling and the analysis techniques used were now available. It is not expected that any statistical uncertainties will change as there will be no new data available due to the LHC run periods (Section 3.1).

The primary updates from ATLAS that would affect the analysis are the change in jet collections from EMtopo (Subsection 7.3.1) to PFlow (Subsection 7.3.2) jets and the updated b-tagging algorithm from MV2c10 (Subsection 7.4.1) to DL1r (Subsection 7.4.2). PFlow jets should help reduce the pile-up, most pronounced with low p_T jets, as such it should most significantly impact the resolved regions. DL1r has better performance than MV2c10 with the largest improvements at high p_T which should help the boosted region in particular. The b-tagging update also mirrors a more general shift in the particle physics community from typical MVA based methods, such as BDTs, to more modern machine learning neural network based techniques, such as DNNs, deep sets and graph networks (Section 6.2). This is apparent in both standard ATLAS techniques (aimed for general use by analyses) and analysis specific techniques (developed for specific use in $t\bar{t}H$).

Overall many things are being re-developed for all the previous signal regions (dilepton, semi-leptonic resolved and semi-leptonic boosted), from top / Higgs reconstruction to signal classifiers. For top / Higgs reconstruction, new Higgs / top taggers were developed for each region with deep-set [109], approaches being used in both the resolved dilepton and single-lepton regions. The Higgs reconstruction for the boosted region, otherwise referred to as boosted-jet tagging, was already DNN based and has been re-trained with the new jet collections and b-tagging updates. A number of studies looking into different network topologies, NN optimisers, extra features (in the form of sub-structure variables), different jet-objects (namely RC-jets Subsection 7.5.2 and LR-jets Subsection 7.5.1) and finally compared to other boosted Higgs-taggers developed centrally by ATLAS for more general use cases. The rest of this section will focus on these studies done for the boosted region.

8.3.1 DNN Model Optimisation

In the initial full Run-2 result the boosted Higgs-tagger was DNN based utilising a simply connected feed-forward DNN with 3 hidden layers each with a width of 100, using the ReLU activation with a softmax final layer giving the probabilities for each category, $P(H)$, $P(t)$ and $P(QCD)$ as an output, such that the sum $P(H) + P(t) + P(QCD) = 1$. It was trained with only ttH data from the MC16a Monte-Carlo campaign which mirrors the ATLAS 2015-2016 data taking periods of Run-2 (see Chapter 5). Because the physics processes being simulated are stochastic in nature, Monte-Carlo simulations are either over- or under- produced compared to the predicted cross-section. They are then either weighted to the correct luminosity ($139 fb^{-1}$ for the full Run-2 data set), this is done either

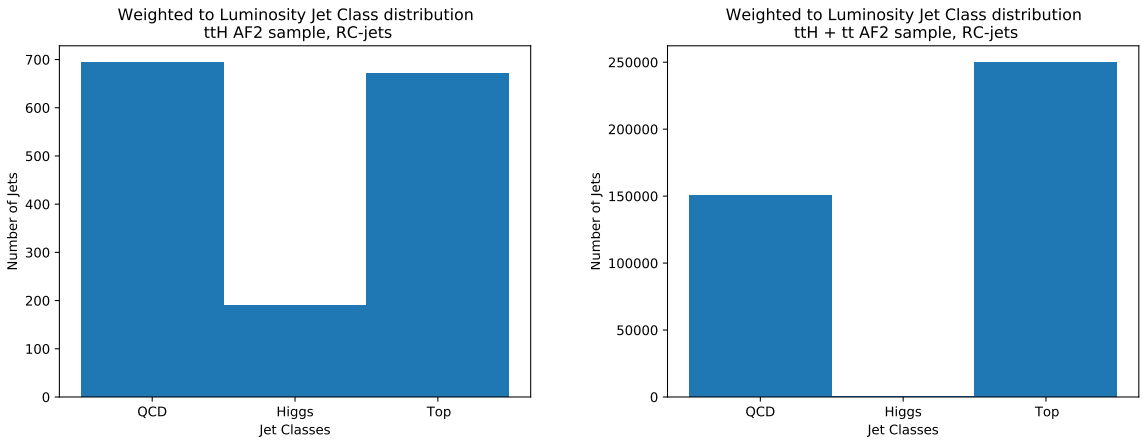
to conserve CPU power in the case of large samples or to ensure adequate statistics in the case of small samples. In the case of $t\bar{t}H$ it is over produced and weighted down to provide adequate statistics.

Acquiring training data is one of the major challenges when training DNNs, because of this the same samples from the analysis were used. To avoid training and testing on the same samples the samples were split into odd and even event numbers, with the DNN trained on odd event numbers being applied to the even event numbers and vice versa. Deep learning typically benefits from training with larger data sets so the full Run-2 MC campaign (MC16a/d/e) was used for training, giving approximately 3 times the training data. Adding in $t\bar{t}$, the primary background, as training data was also investigated. However, this only further unbalances the data set for training (see Figure 8.35b). For classifiers the training data, for optimal results, should be equally balanced between classes. This can be achieved in two ways by either weighting the loss or randomly oversampling certain classes, with the former being chosen.

The class imbalance can be seen from Figure 8.35, which must be accounted for during training as the network should "see" balanced classes. Otherwise for a correct guess the network can simply pick the dominant class as it will be a correct guess the majority of the time. MC samples are either over/under produced compared to the expected cross-section for a given luminosity and then weighted to the expected cross-section. This is done in order to either provide adequate statistical variation for small samples, such as $t\bar{t}H$, or to use less CPU cycles for large ones, such as the $t\bar{t}$+jets background. These normalisation weights were then built into the class weights used in training with

$$(W_{QCD}, W_{Higgs}, W_{top}) = \frac{\sum_n^N Wnorm_n}{3 \times (\sum_q^{N_q} Wnorm_q, \sum_h^{N_h} Wnorm_h, \sum_t^{N_t} Wnorm_t)},$$

where the sum of these normalisation weights gives the total number of jets scaled to luminosity with the subscript n denoting for all jet classes and q, h, t for Higgs,top and QCD classes respectively. With this scheme it was decided that adding the $t\bar{t}$ background sample into training was not desirable because it would reduce the normalised Higgs fraction too extremely (see Figure 8.35b) and cause rounding errors in the class weights such that the classes would not be properly balanced.

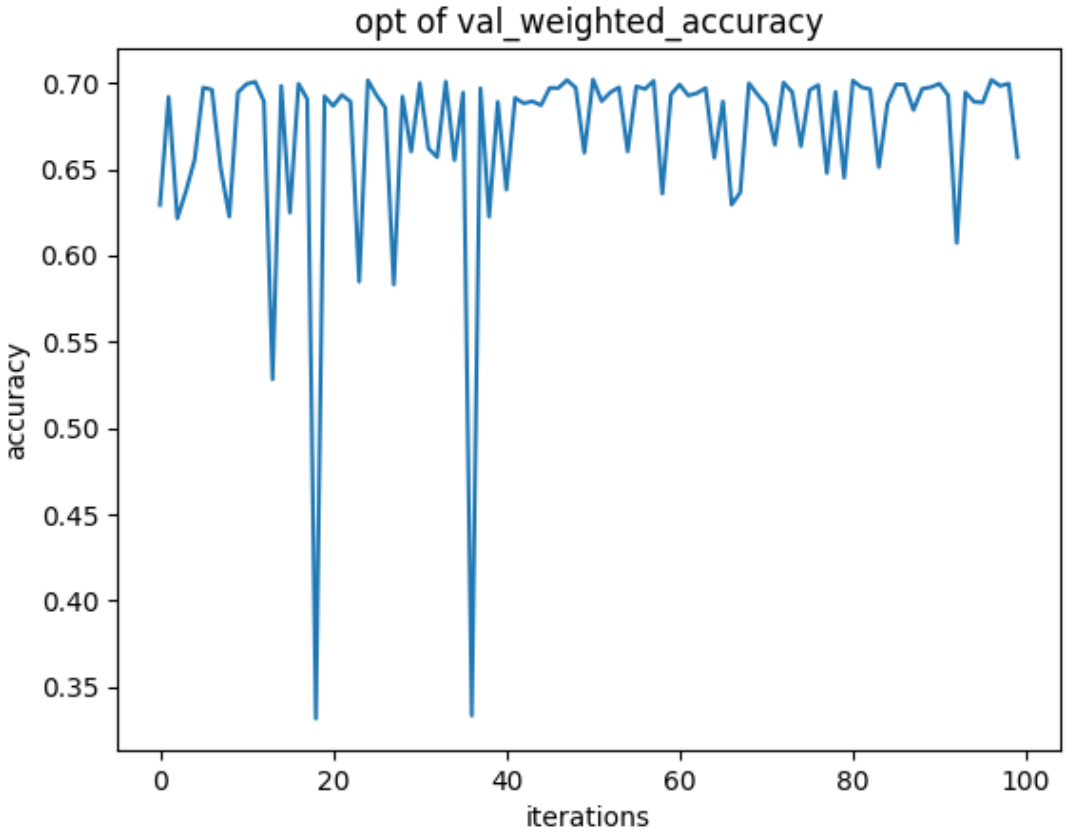


(a) Class distribution of jets in the $t\bar{t}H$ sample weighted to luminosity.

(b) Class distribution of jets in the $t\bar{t}H + t\bar{t}$ samples weighted to luminosity.

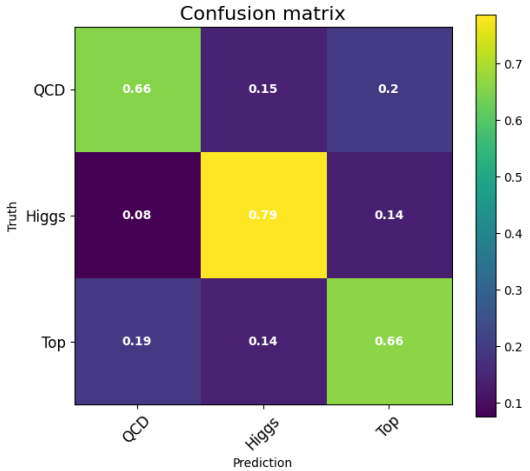
Figure 8.35: Class distributions for $t\bar{t}H$ and $t\bar{t}$, from the AF2 sample using RC-jets.

While generating the new samples the hyper parameters of the network were tested with the Bayesian optimisation algorithm described in Subsection 6.2.5 to see how optimal they were. However, it is often the case that this approach can yield little to no results thus it was only perused while the new samples were being generated. An implementation from the Scikit-Optimize python library was used, with the surrogate function being a Gaussian Process and using the default gp_hedge acquisition function. The search space used was: the width of hidden layers [50,100], the number of hidden layers [3,5], the optimisation algorithm being used by the DNN [adam, stochastic-gradient-decent (SGD)] (typically the popular adam algorithm converges faster than SGD), the learning rates [0.01:0.05] (motivated by the modified class weighting). Some input parameters were also added to the search space namely the substructure variables: d_{12}, d_{23}, Q_w, the subjets inv.-mass of b-jets and the subjets inv.-mass of $light$-jets. This was done because it was observed from the previous analysis that these were the worst modelled distributions in terms of data/MC agreement. The metric being minimised was $1-accuracy$ (see Subsection 6.2.5) which was weighted with the class weights. The results of the hyper parameter optimisation can be seen in Figures 8.36 with no real performance gain found. The iterations of the algorithm did not surpass an accuracy of 0.70 (see Figure 8.36a), with its best variation using all of the input features, a depth of 3, a width of 84, a learning rate of 0.01 and the adam optimizer. The training history of the "best" model can be seen in Figure 8.36b and its confusion matrix in Figure 8.36c. However, this performance was only marginally better than the setup from the previous round which had an accuracy of 0.68. This is a small enough difference to be accounted for by statistical variations between trainings and so it can be concluded that the previous setup was already optimal.



(a)

(b)



(c)

Figure 8.36: Overall performance of Bayesian hyper-parameter optimisation iterations evaluated with the accuracy metric (a). Training history (accuracy) of the best performing model (b) and its confusion matrix (c).

It was found that the popular adam optimiser (described in Section 6.2) converged significantly faster so variations using both the adam and SGD optimisers were trained

with the new samples. However, the variations trained with the adam optimiser started to exhibit over-training at around 50 epochs where the training and validation sets started to diverge in metrics such as accuracy and precision (see Figure 8.38), so a training epoch of 50 was chosen. It was found that the SGD optimiser variations would exhibit larger variations between successive trainings compared to the previous jet collection, this can be seen from the training histories in Figure 8.37.

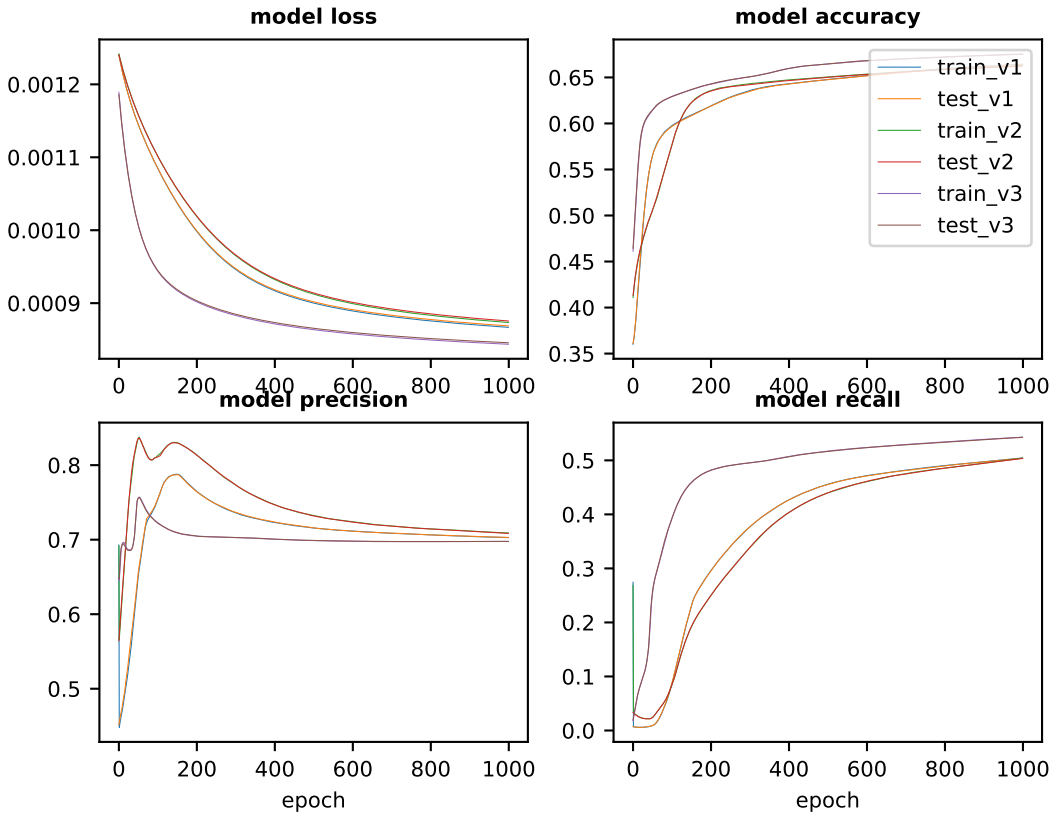


Figure 8.37: Repeated training histories to evaluate performance and reproducibility of the SGD optimiser upon re-training the network 3 times via metrics (loss, accuracy, precision and recall) with Pflow-jets.

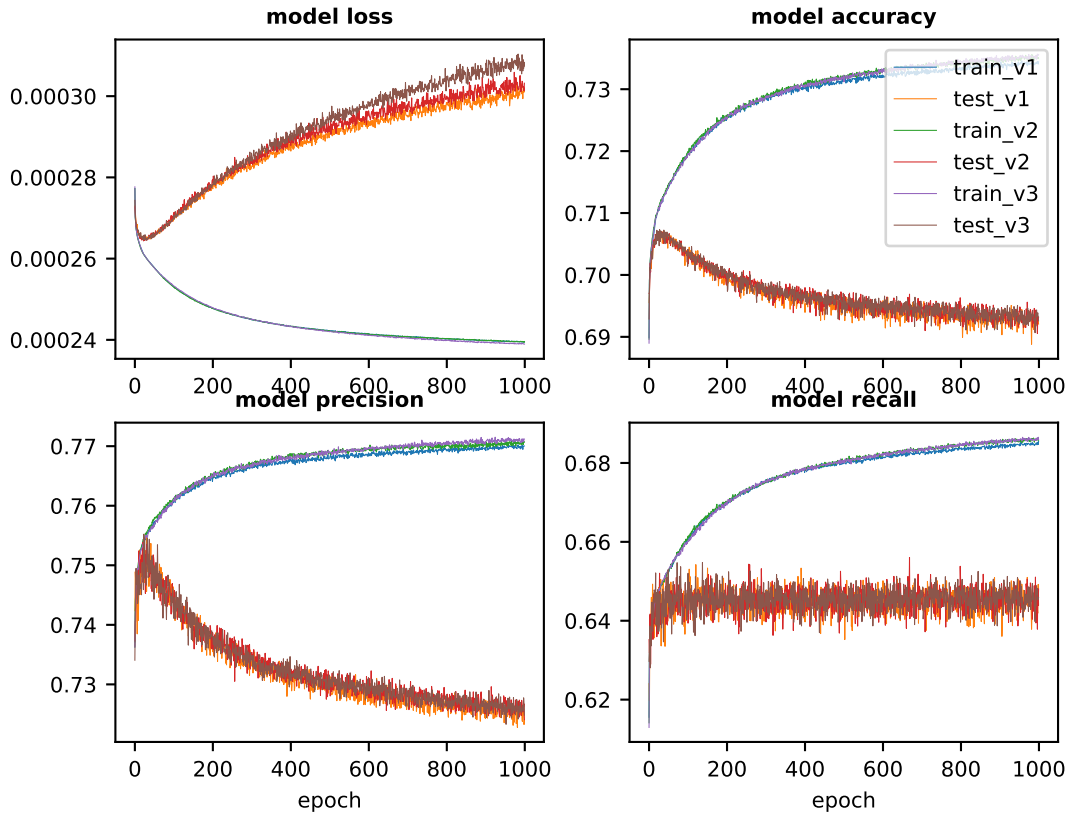


Figure 8.38: Repeated training histories to evaluate performance and reproducibility of the adam optimiser upon re-training the network 3 times via metrics (loss, accuracy, precision and recall) with Pflow-jets.

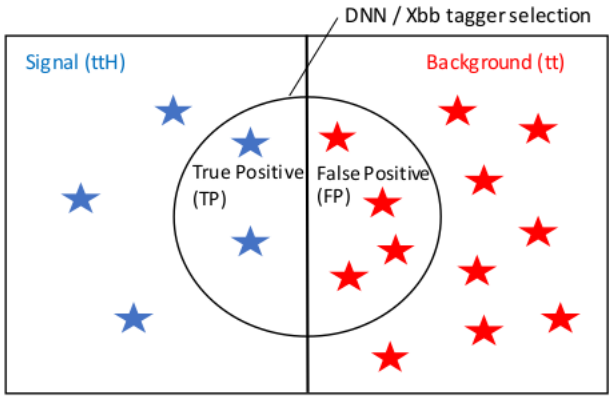


Figure 8.39: Metrics, purity = selected $t\bar{t}H/t\bar{t} = TP/FP$, $t\bar{t}H$ selection efficiency = $TP/t\bar{t}H$ otherwise known as true positive rate.

The DNN variations were then compared with more analysis specific metrics, namely the selected $t\bar{t}H/t\bar{t}$ referred to as purity, an approximation of s/b since $b > t\bar{t}$ it wasn't refereed to as such, and the $t\bar{t}H$ selection efficiency as can be seen in Figure 8.39. When comparing the $P(H)$ distributions in a loose region selecting an event with ≥ 1 RC-jet

with ≥ 2 subjets. It can be seen that the SGD variation has a higher selection efficiency (0.484 vs 0.369) when selecting events with at least 1 Higgs tag, defined here as an RC-jet with $P(H) > P(t)$ and $P(H) > P(QCD)$. This can be seen in Figures 8.40c and 8.40d although the adam variation achieves a higher purity this can be adjusted with the cut on $P(H)$. It should also be noted that due to the large variations found with the SGD optimiser between successive trainings (see Figure 8.37) the $P(H)$ distribution (Figure 8.40a) is not particularly smooth. This is due to the final DNN being trained/validated on odd/even samples (and vice versa) and then stitched together.

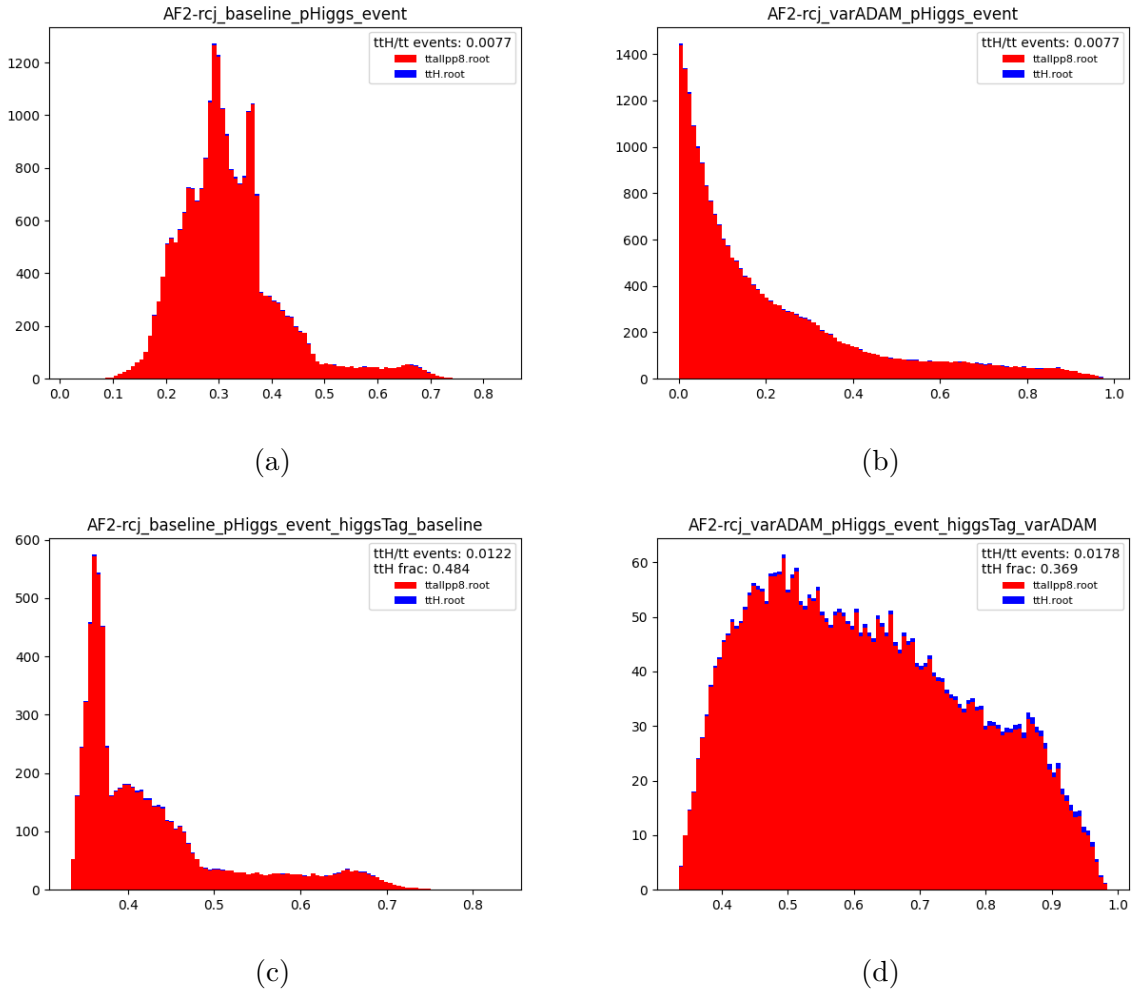
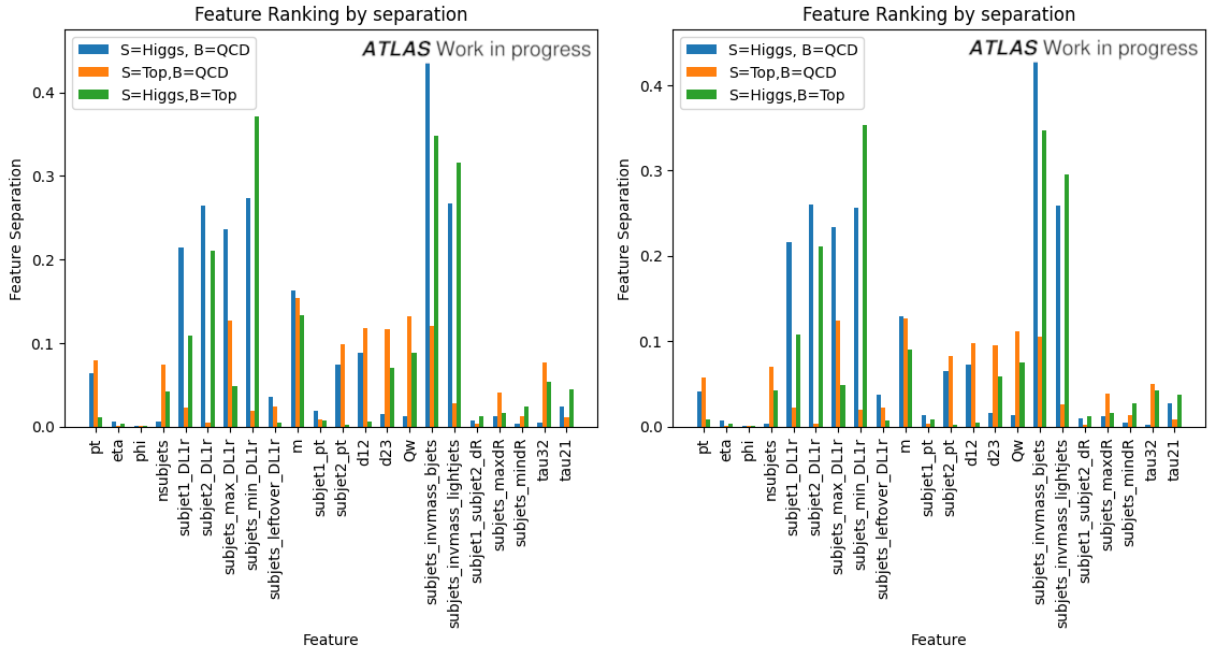


Figure 8.40: The Higgs probability $P(H)$ DNN output distribution comparison for SGD (a) and adam (b) DNN variations and their respective Higgs tagged events (c) and (d) for AF2 $t\bar{t}H$ events using RC-jets.

Because of this the SGD algorithm was chosen for the next round of studies. These included investigating the effectiveness of different boosted jet collections, namely RC-jets and LR-jets described in Subsection 7.5.2 and Subsection 7.5.1, and if using additional substructure variables, namely the n-subjetyness ratios τ_{32} and τ_{21} (described in Subsection 7.5.3), as features during training provides any additional benefit to the network. Because

LR-jets are not re-clustered from small-r jets it is less trivial to associate small-r jets with them. This is required because the ATLAS b-tagging algorithms are designed to work with small-r jets and seeing as the $H \rightarrow bb$ decay mode is being investigated, matching two b-tagged jets to a boosted object is one of the primary ways to identify a boosted Higgs decay. This sort of substructure information, among others, is one of the reasons that feature engineering (see Subsection 6.2.6) is used to try to improve the performance of the network. The matching scheme used here to match small-r jets to LR-jets was simply ΔR matching described in Subsection 7.5.1, as opposed to using the more complicated ghost-track jets. Then the various training features used for both sets of jet collections were evaluated with separation power in terms of $Higgs$, top and QCD jets this can be seen in Figure 8.41. It can be seen that for some features the separation power is more pronounced with RC-jets as is the case with the n-subjetyness ratios τ_{32} and τ_{21}. The use of certain features and jet-collections requires the full simulation of the ATLAS detector as opposed to the fast simulation algorithm (AF2). The details of this are laid out in Section 5.5, this is a key consideration when assessing the performance or potential modelling benefits that must be kept in mind when designing the analysis as full simulation using GEANT-4 [110] can be rather CPU intensive.



(a) Separation power of input features with RC-jets.

(b) Separation power of input features with LR-jets.

Figure 8.41: Separation power of input features between $Higgs$, top and QCD jets.

With the above mentioned in mind 4 new DNNs were trained using SGD and full simulation and assessed, with $t\bar{t}H$ purity and $t\bar{t}H$ fraction to see if there was any benefit, to including these features during training. Both LR-jets and the n-subjetyness ratios

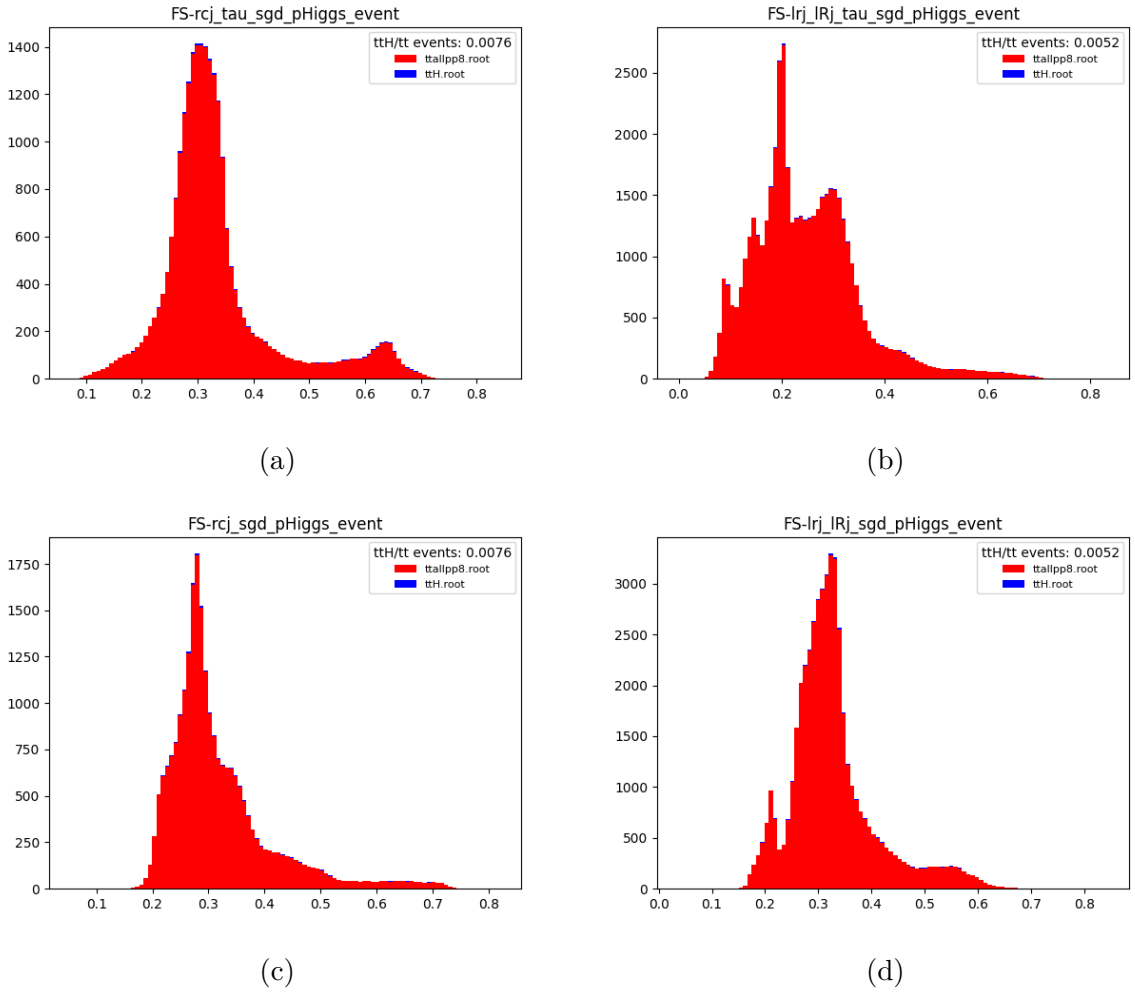


Figure 8.42: The Higgs probability $P(H)$ DNN output distribution comparisons of RC-jets (a,c) and LR-jets (b,d) trained with additional n-subjetyness variables, τ_{32} and τ_{21}, (a,b) for FS $t\bar{t}H$ events.

require full detector simulation. The output distributions can be seen in Figures 8.42d with the additional τ variables creating an interesting looking hump in both jet collections at around 0.65 in Figure 8.42a and 0.35 in Figure 8.42b. This is indicative of the lower separation power achieved in these τ variables with the LR-jets (see Figure 8.41).

However, after selecting events with a Higgs tagged boosted-jet, it can be seen that there was no performance gain from adding the τ variables. This can be seen in Figure 8.43, with the RC-jet DNN variations achieving the best selection efficiency. There is no notable difference in selection efficiency or sample purity between the RC-jet DNNs trained with or without the τ variables. A final comparison of these two metrics (selection efficiency and sample purity) for the DNN variations can be seen in Figure 8.44 which shows the RC-jet DNN variations performing better (achieving a higher sample purity for a given selection efficiency) than the LR-jet variations, with similar performance observed between RC-jet variations. Because of this the 3 DNNs using RC-jets were selected for

further comparisons with the centrally available Xbb tagger using LR-jets. The training histories for the final 3 variations can be seen in Figures 8.45 to 8.47 each trained for 200 epochs with the precision of both odd and even variations converging. Even though the LR-jets and τ variables did not provide a reason to request full simulation samples they also did not provide a reason not to request full simulation studies as the main benefit for using full simulation would become apparent in the Data/MC agreement of the DNN input features.

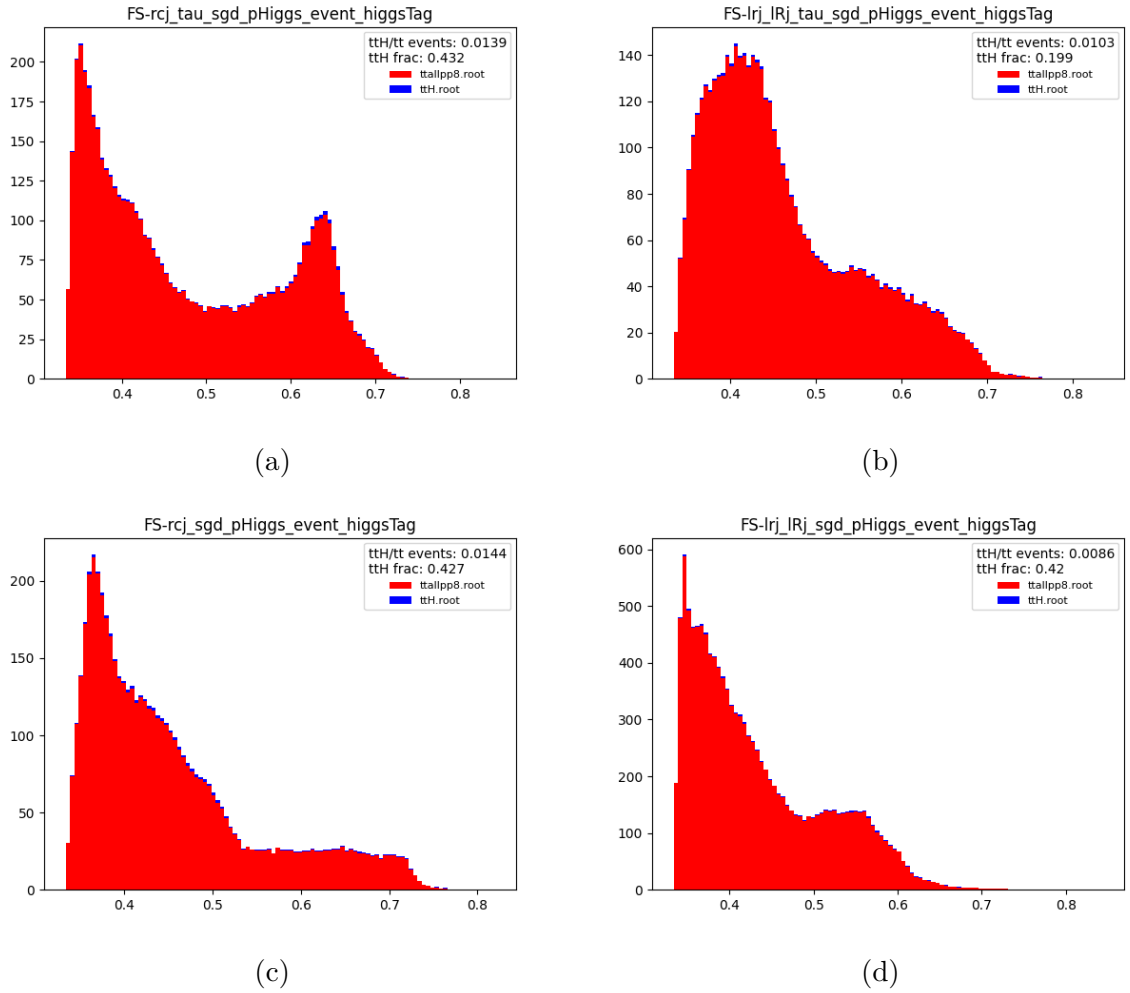


Figure 8.43: The corresponding Higgs tagged events for Figure 8.42 comparing the Higgs probability $P(H)$ DNN output distributions for RC-jets (a,c) and LR-jets (b,d) trained with additional n-subjetyness variables, τ_{32} and τ_{21}, (a,b) for FS $t\bar{t}H$ events.

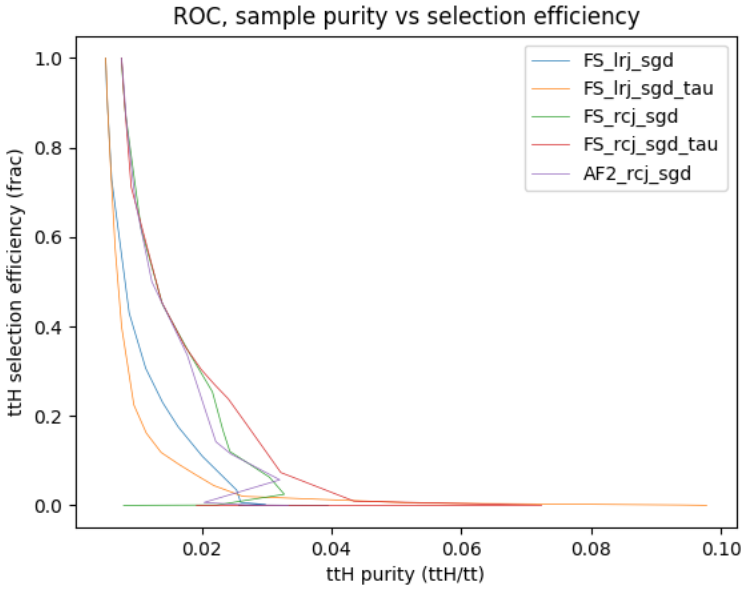


Figure 8.44: ROC curve between $t\bar{t}h$ selection efficiency and $t\bar{t}H$ purity for DNN variations comparing: FS samples to AF2 samples with RC-jets and RC-jets to LR-jets with FS-samples.

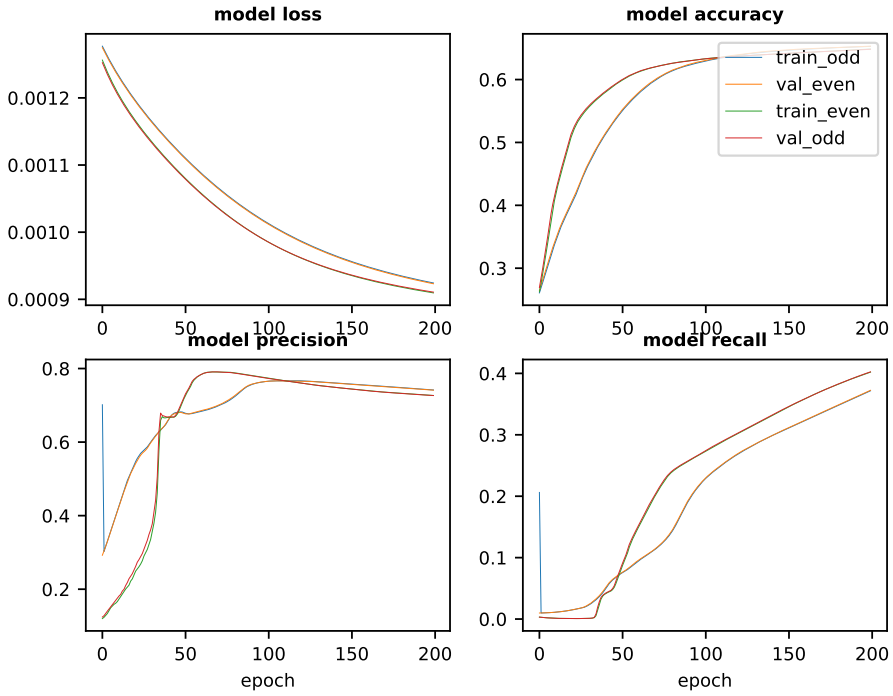


Figure 8.45: Training history metrics (loss, accuracy, precision and recall) for the AF2 model using RC-jets.

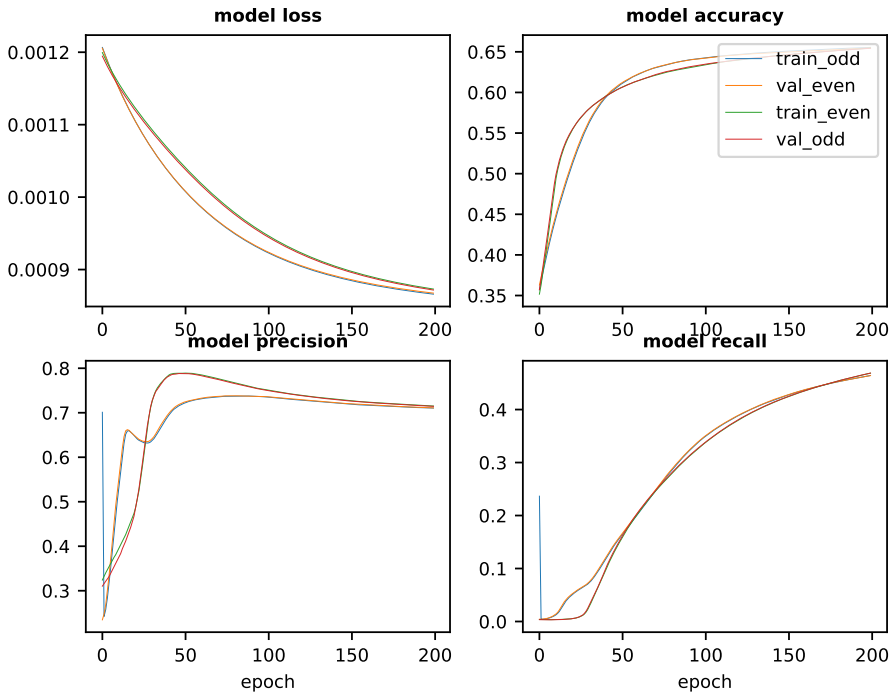


Figure 8.46: Training history metrics (loss, accuracy, precision and recall) for the FS model using RC-jets.

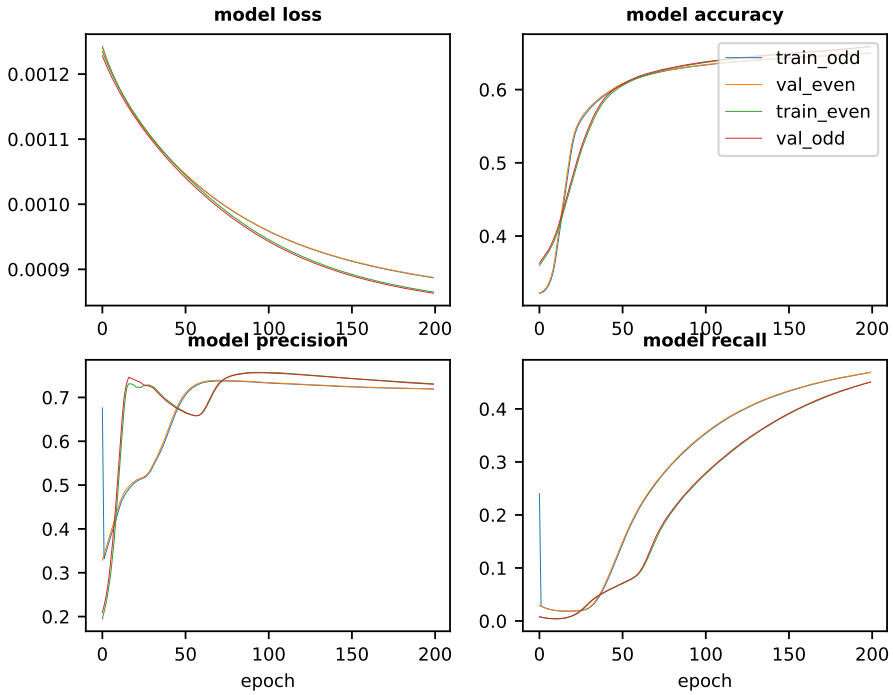


Figure 8.47: Training history metrics (loss, accuracy, precision and recall) for the FS model with additional n-subjetyness variables using RC-jets.

8.3.2 Xbb tagger studies

Taking the most successful DNN variations (using RC-jets) comparisons are made with the centrally produced Xbb Higgs tagger. This is done in two regions: a loose region to assess the overall performance of jet tagging; and the boosted signal region, from the previous analysis, to assess the performance in the region most sensitive to the signal. These regions are defined below:

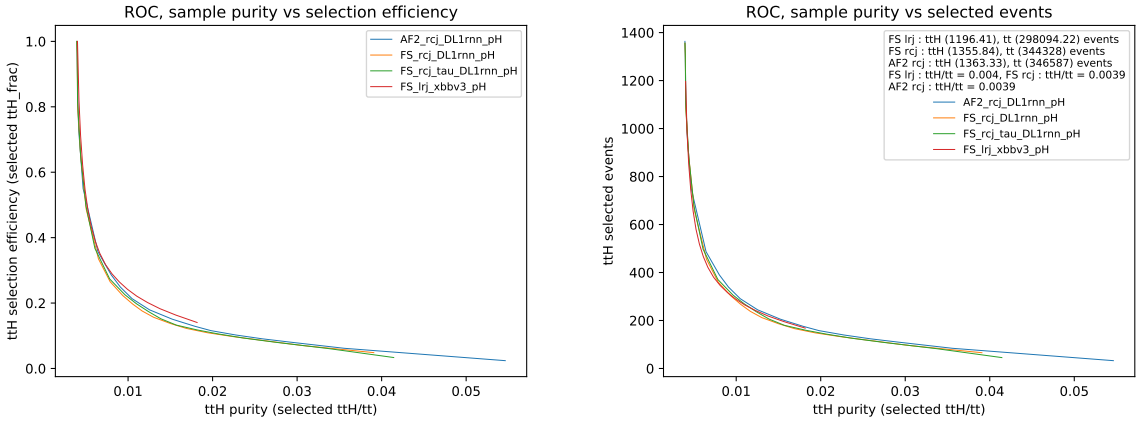
- Loose region: ≥ 4 jets & ≥ 1 RC-/LR-jet with ≥ 2 sub/ΔR-jets.
- Boosted region: ≥ 4 jets & (≥ 4 b-tag @ 85 & ≥ 2 b-tag @ 77 jets) & (≥ 1 RC-/LR-jet with $p_T \geq 300$ GeV & and ≥ 1 RC-/LR-jet with ≥ 2 sub/ΔR-jets) ≥ 1 RC-/LR-jet with $100GeV \leq M < 140$ GeV and $= 2$ b-tag subjets @ 85 + ≥ 2 additional b-tag jets @ 77.

The main advantages of utilising this tagger would be that it was trained with a different data set, so it could potentially have a larger training set as well as remove potential sources of human error e.g training and running the network on the same subset of data. It was trained using LR-jets p_T and η information and the 3 leading sub-jets b-tagging information in this case the outputs of the DL1r algorithm ($P(b)$, $P(c)$, $P(light)$) with the LR-jets "sub-jets" being defined with ghost associated track-jets scheme described in Section 7.5.1. An RS graviton $\rightarrow hh \rightarrow bbbb$ was used as the signal sample and JZW dijet + $Z' \rightarrow tt$ was used as the background sample. Because the Xbb tagger was designed as a general purpose tagger, intended for use both in searches e.g Higgs doublet models and for SM measurements, there was a large emphasis put on decorrelating it from the invariant mass of the Higgs and so no sub-structure variables or any invariant mass quantities were used. Due to this only the LR-jet kinematics and the associated small-r jets b-tagging information was used, as such feature engineering was not utilised to increase the performance. Rather it was hopped that given a deep enough network with a large enough data set it would be able to learn specific features from the data. It can be seen from Figure 8.41 that the sub-jet type b-tagging information are the most important in terms of Higgs/top separation (in green) with the boosted-jets invariant mass followed by Q_w which is the invariant mass of the W boson and indicative of a top decay. The invariant mass of the sub-jets *light*-jets are still counted as b-tagging information since it uses the DL1r algorithm to tag them as such. So overall apart from invariant mass quantities the Xbb tagger is supplied with much the same information as the DNNs specifically trained for $t\bar{t}H$. The Xbb tagger also gives 3 output probabilities for each class $P(H)$, $P(top)$ and $P(light)$, these are then combined into a discriminant, $D = \ln\left(\frac{P(H)}{(1-f_{top})P(QCD)+f_{top}P(t)}\right)$, with f_{top} being set to 0.25.

The main reason for including comparisons in the Loose region was to remove any bias from ΔR matching small-r jets with LR-jets being introduced into the region definition

prior to comparing the taggers themselves. The taggers were then assessed with similar ROC curves comparing selection efficiency and sample purity described previously in Subsection 8.3.1. This was done both with the $P(H)$ output and with the discriminant D, with the DNNs definition mirroring the Xbb definition.

From the ROC curves in Figure 8.48 comparing the performance in the Loose region we can see that the Xbb tagger performs comparably to the DNNs in both relative terms (Figures 8.48a) and on a per event biases (Figure 8.48b). When looking at our sample purity vs selection efficiency ROC (Figure 8.48a) there is little distinction to be made however, the Xbb curve (red) fails to achieve a sample purity greater than 0.2 although produces a slightly higher selection efficiency. The difference in selection efficiency disappears when looking at the raw number of events numbers as can be seen in Figure 8.48b and is due to the slight lower number of events with LR-jets. It can also be seen that if selecting a tight cut then the DNN variations can achieve a selection with considerably higher sample purity. The story is much the same when looking at the standard ROC curve of true positive rate (TPR), in this case selection efficiency, and false positive rate (FPR). Because they are inconclusive they have been included in Appendix D in Figures D.1a and D.1b. Again the Xbb tagger performs comparably when looking at relative events and slightly worse on raw events.



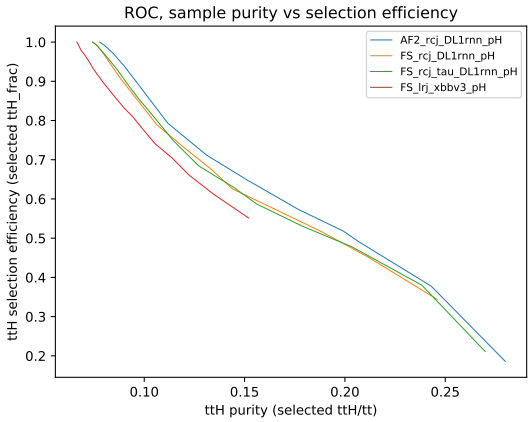
(a) Selection efficiency (relative) vs $t\bar{t}H$ purity on an event basis.

(b) Selection efficiency (absolute) vs $t\bar{t}H$ purity on an event basis.

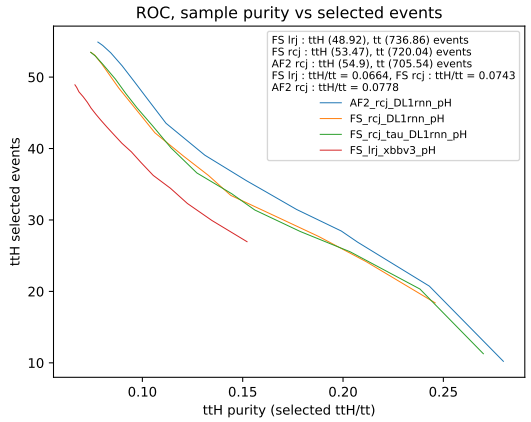
Figure 8.48: ROC curves comparing DNN variations and the Xbb tagger in the Loose region.

When comparing them in the boosted region, they are both compared to each other as well as the results from the previous paper, where a DNN cut of $P(H) = 0.6$ was used. As these new DNNs have been re-trained a DNN cut of $P(H) = 0.6$ used previously will not behave the same and rather an appropriate cut should be chosen with similar performance. This was done by selecting $P(H)$ values with a similar number of $t\bar{t}$ background

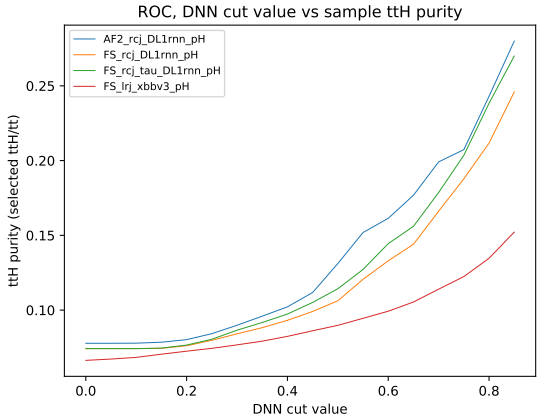
events. For the ROC curves in Figure 8.49 the difference between the Xbb tagger and the DNN variations are more pronounced showing clearly better performance with the DNN variations in both TPR (selection efficiency) and sample purity.



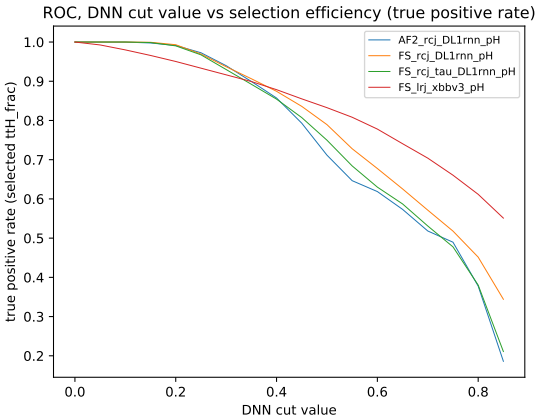
(a) Selection efficiency (relative) vs $t\bar{t}H$ purity on an event basis.



(b) Selection efficiency (absolute) vs $t\bar{t}H$ purity on an event basis.



(c) Cut values vs $t\bar{t}H$ purity on an event basis.



(d) Cut values vs TPR / selection efficiency (relative) on an event basis.

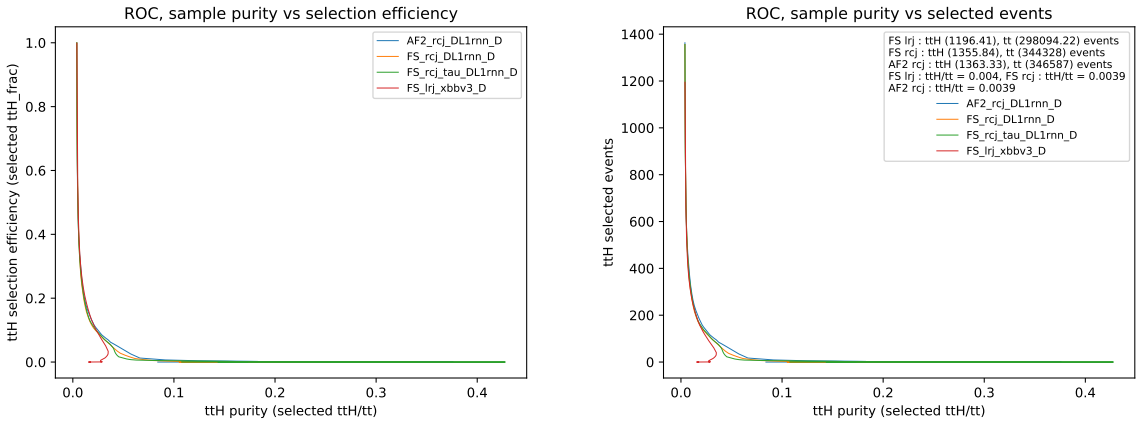
Figure 8.49: ROC curves comparing DNN variations and the Xbb tagger in the Boosted region.

From Table 8.7 we can see that when selecting for a similar (or less) number of background events the DNNs perform better with a smaller number of background events (460.6 for AF2) and a higher number of signal events (47.1 for AF2) using a $P(H)$ cut of 0.4 for the AF2 variation. This shows promising performance with a slight increase compared to the last round.

DNN var.	DNN cut	$t\bar{t}H$ purity	$t\bar{t}H$ sel. eff.	$t\bar{t}H$	$t\bar{t}$	*Data*
paper	0.6	0.090	-	43.6	483	699
AF2	0.40	0.102	0.86	47.1	460.6	-
FS	0.45	0.099	0.84	44.7	451.0	-
FS + τ	0.40	0.097	0.85	45.7	469.1	-

Table 8.7: Comparison of DNN cuts with the previous analysis. The DNN variations have a higher $t\bar{t}H$ signal yield when selecting for a similar number of $t\bar{t}$ background samples.

However, when comparing the Xbb tagger to the DNN variations it should be noted that as it uses LR-jets it has to be calibrated as opposed to using a DNN trained with RC-jets whos individual components are already calibrated. This is because RC-jets are re-clustered from small-r jets which are already calibrated. Because of this the Xbb tagger has to be used with one of its pre-defined W.P. which are defined with the discriminant $D = 3.09$ (50%), $D = 2.44$ (60%), $D = 1.76$ (70%). As such in Figures 8.50 8.51, the previous ROC curve comparisons are repeated using the discriminant D and tell the same story as before. The artifacting that can be seen in Figure 8.51 is due to the discriminant D being undefined for some points at low selection efficiencies.

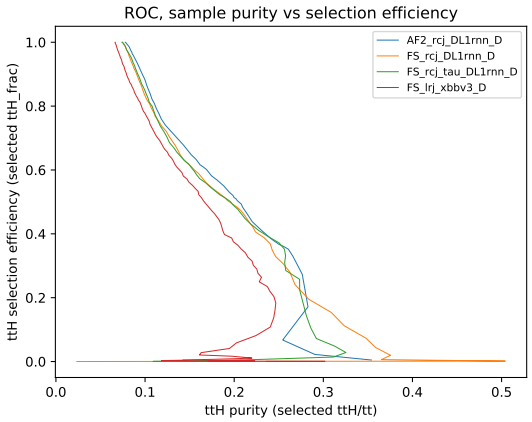


(a) Selection efficiency (relative) vs $t\bar{t}H$ purity on an event basis.

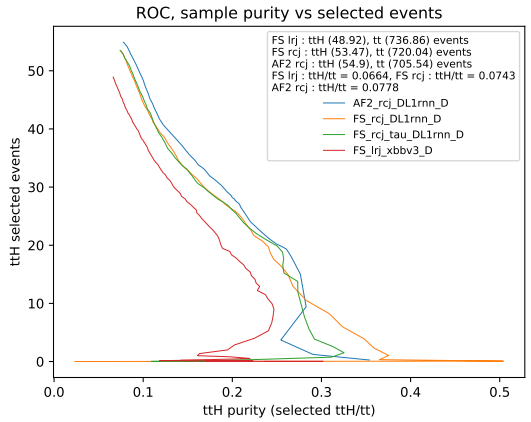
(b) Selection efficiency (absolute) vs $t\bar{t}H$ purity on an event basis.

Figure 8.50: ROC curves comparing DNN variations and the Xbb tagger using the discriminant D in the Loose region.

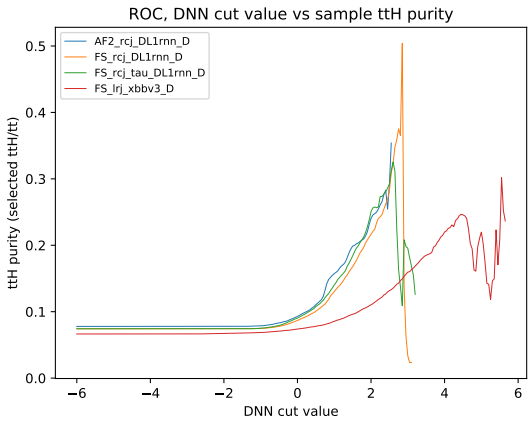
Because the Xbb taggers W.P. must be used a comparison of all 3 W.P. is outlined in Table 8.8 with DNN D values selected with similar number of signal events for each. It can be seen that for each of the Xbb W.P. the DNNs select a lower number of $t\bar{t}$ background events and achieve a higher sample purity.



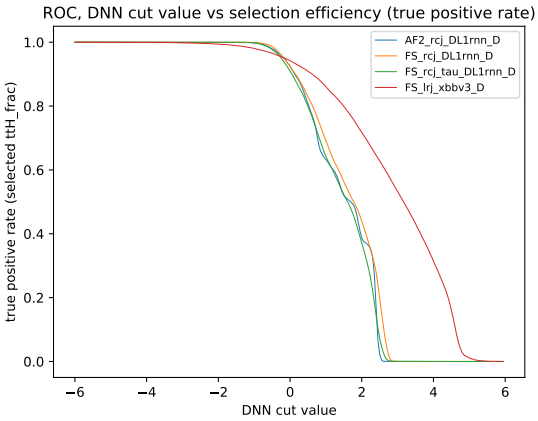
(a) Selection efficiency (relative) vs $t\bar{t}H$ purity on an event basis.



(b) Selection efficiency (absolute) vs $t\bar{t}H$ purity on an event basis.



(c) Cut values vs $t\bar{t}H$ purity on an event basis.



(d) Cut values vs TPR / selection efficiency (relative) on an event basis.

Figure 8.51: ROC curves comparing DNN variations and the Xbb tagger using the discriminant D in the Boosted region.

DNN	D	$t\bar{t}H$ purity	$t\bar{t}H$ selection efficiency	$t\bar{t}H$	$t\bar{t}$
Xbb	3.10	0.163	0.51	25.1	153.8
Xbb	2.45	0.130	0.64	31.1	239.8
Xbb	1.75	0.103	0.76	37.2	361.0
AF2	1.85	0.214	0.47	25.5	119.3
FS	1.90	0.208	0.47	25.2	121.1
FS + τ	1.70	0.204	0.48	25.5	125.1
AF2	1.35	0.181	0.57	31.2	172.2
FS	1.40	0.160	0.59	31.5	196.7
FS + τ	1.25	0.158	0.58	31.3	197.9
AF2	0.80	0.143	0.68	37.3	261.4
FS	0.95	0.127	0.71	37.7	297.5
FS + τ	0.80	0.124	0.70	36.6	302.2

Table 8.8: Comparison of the DNN- with the Xbb-taggers. The DNNs have a lower $t\bar{t}$ background yield when selecting for a similar number of $t\bar{t}H$ signal samples.

Overall from these studies there is no observed benefit from using full simulation samples and the extra τ n-subjetyness variables as training features. The Xbb tagger also performs worse when looking in the Boosted Signal region as defined with the previous analysis. For a similar performance to the previous paper a recommendation for using the AF2 variation, with a $P(H) = 0.4$ cut, can be made as it provides the best performance when compared to the previous paper, the other DNNs and the Xbb tagger selecting 47.1 $t\bar{t}H$ events with less background events and achieving a higher sample purity. In addition it requires no full simulation samples. However, if full simulation samples are requested, motivated by other studies, than it would be advantages to also use one of the DNNs trained with those samples selecting a cut value from Table 8.7. It can be assumed that the Xbb tagger performs worse than bespoke DNNs trained here because the features used in the DNNs were specifically chosen for Higgs / top separation, the Xbb tagger has been trained with a smaller top quark fraction and was designed to not make use of any invariant mass quantities.

8.4 Summary

Overall there is promising performance shown for boosted topologies in $t\bar{t}H$ analyses with significant contributions when combined with resolved topologies.

In the all-hadronic feasibility study in Section 8.1, it was shown that due to the high QCD-multijet background boosted topologies suffer with low statistical significance. It is assumed that this is because of the presence of real b-jets coming from the QCD-multijet background that can't discriminated against with b-tagging algorithms such as MV2c10.

In the Lepton + Jets Full Run-2 Analysis in Section 8.2, the contribution of the boosted region to the final measurement was shown. This includes its contributions to both the inclusive fit and the differential cross-section (STXS) measurements. This was overall significant and was extremely important for the final two STXS bins $[300, 450)$ and $[450, \infty)$.

Finally in Section 8.3, on-going work re-developing the boosted DNN tagger for the Lepton + Jets Legacy Analysis is shown. From this it can be seen that bespoke DNN-Higgs taggers still outperform ones aimed for more general purposes. A slight performance improvement can also be observed when comparing the newly trained DNN to the previously trained one used in the Lepton + Jets Full Run-2 Analysis.

Chapter 9

Conclusions

Measurements of the $t\bar{t}H$ cross-section are the most direct way of accessing the top Yukawa coupling (y_t), one of the free parameters of the SM. Since the $H \to b\bar{b}$ decay mode is the most prevalent it stands to reason for analyses to specifically target this decay, in order to gain as much statistics as possible, given the already rare nature of $t\bar{t}H$ production at the LHC.

Overall the work presented here was focused on applying ML techniques for boosted Higgs reconstruction, with the $H \to b\bar{b}$ decay mode, for use in $t\bar{t}H$ analyses. There were three separate studies presented:

- An all-hadronic feasibility study, which aimed to assess the statistical significance of boosted topologies;
- Lepton + Jets Full Run-2 Analysis, which showed the contributions provided by the semileptonic boosted region to the final result;
- Lepton + Jets Legacy Analysis, which described the retraining of the deep neural network (DNN) used for boosted Higgs identification in the upcoming paper.

From the all-hadronic feasibility study it was seen that boosted topologies suffered with low statistical significance. This was assumed to be because of real b-jets present in the QCD-multijet background which are not rejected by b-tagging algorithms. This intern led to the QCD-multijet background being irreducible thus making any contributions statistically insignificant.

The Lepton + Jets Full Run-2 Analysis uses $139fb^{-1}$ of data at $\sqrt{s} = 13$ TeV. It is the first $t\bar{t}H$, $H \to b\bar{b}$ differential cross-section measurement by ATLAS. It can be seen that the use of DNNs for boosted Higgs reconstruction yielded impressive results when compared to the BDTs used in the resolved regions, see Figure 8.12. The result of the inclusive profile-likelihood fit is $\mu = 0.35^{+0.36}_{-0.34}$ which corresponds to $\sigma = 1.0(2.7)$ observed(expected) significance compared to the background-only hypothesis. Overall the

result is compatible with the previous $36.1 fb^{-1}$ ATLAS result [23], which used a subset of the data, with the impact of the systematics being reduced by a factor of two. The boosted region was primarily defined through the selection of a boosted Higgs candidate in the event, which was identified with the DNN tagger. The DNN primarily made use of b-tagging, ΔR and kinematic information as well as some substructure variables that do not use the clustering history of the jet algorithms. The contribution of the boosted region to the inclusive result as well as to the differential cross-section, STXS bins $[300, 450)$ and $[450, \infty)$ GeV, are significant $\mu = 0.32^{+0.61}_{-0.57}$.

The Lepton + Jets Legacy Analysis is a re-analysis of the Run-2 ATLAS dataset with updated jet collections, b-tagging algorithms, and analysis techniques. It is also set to benefit from improved $t\bar{t}$ modelling. The retraining of the DNN for boosted Higgs classification is shown, along with comparisons made between different boosted jet collections, including re-clustered/large radius jets. It was found that the previously used re-clustered jets were the more advantageous option due to their ease of region combination with resolved regions (since systematics directly propagate over), the fact that they don't require full detector simulation and can be used with Atlfast-2 samples, and that they overall have smaller uncertainties than large radius jets. It was also shown that training bespoke DNNs for boosted Higgs reconstruction was still advantageous for the analysis, especially as the general-purpose boosted $H \to b\bar{b}$ (Xbb) taggers available in ATLAS are only trained for large radius jets. Additionally, the heavy focus on mass decorrelation necessitated by the general intended use of the Xbb tagger resulted in worse Higgs/top separation, resulting in a lower yield of $t\bar{t}H$ events with a higher background contamination when applied to the previously used signal region. As such, the advantages of the bespoke DNN are numerous enough to warrant the additional time taken to train it.

With ATLAS having commenced taking data in Run-3, there will be ample statistics available for improved measurements of $t\bar{t}H$, $H \to b\bar{b}$. The anticipated Run-2 Legacy result, currently awaiting unblinding, also benefits from considerable improvements to both resolved regions, with DNN-based techniques now reconstructing the Higgs candidates in the respective regions. The boosted and resolved semileptonic regions now share a common deep set for event classification, which can help with edge cases where one event may be better reconstructed in the resolved or boosted regions. There are also early indications of improvements in $t\bar{t}$+jets modelling when comparing to the previous result, which was dominated by $t\bar{t}$+jets modelling systematics.

Appendix A

Qualification Task ReadMe

This appendix shows the ReadMe file made for the metrology script designed for the ITk strip production, which served as my qualification task, detailed in Chapter 4. The ReadMe served as the sign-off document for the task and details how to draw arbitrary geometries of chips with the software package.

METRO

This package provides a framework to perform checks on module metrology in the x-y plane. It is intended to be used for the ATLAS ITK strip modules upgrade and will work for both the barrel and end-cap modules. It primarily calculates the maximum and minimum wire bonding angles for hybrids given the measurement of 2 of the fiducials on the hybrid. It also will calculate if there are any physical constraints (i.e. the hybrid is overhanging the module). It then outputs the results into a JSON file.

Description

It is written in Python 2 and uses mostly standard python packages:

- copy
- json

As well as common non-standard packages:

- matplotlib
- numpy

Usage

To run:

```
./main.py
```

The strip-modules will have to be added to the modules python package and added to the nested if statement at the begging of the setup function in `setup.py` i.e.

```
elif str_module == "ls":
from modules import ls as cfg
```

The modules are defined as dictionaries in the modules directory and should be added as `strip-module.py` (i.e. `ls.py`)

```
touch modules/ls.py
```

Then a line should be added to the `modules/__init__.py` file importing the dictionary i.e.

```
from ls import ls
```

A module dictionary is defined with the a coordinate system that has the origin at the bottom left corner. All objects on the module are defined in this way and their relative bottom left corners (blc) then define their coordinates on the module. Objects are to be defined as a set of coordinates going anti-clockwise starting from it's bottom left corner. Objects can be arrayed but the spacing must be specified.

The allowable objects that can be defined are:

- module
- hybrids
- ASICs
- strip pad
- constraints (i.e. power board)

At the top level of the dictionary the module is defined with the keys:

- `module` - containing it's name
- `def` - the list of x,y coordinates given anti-clockwise
- `"str"_hybrid` - where "str" is a unique string
- `strip_pad`
- `bond_angle` - giving the minimum wire bonding angle
- `"str"_constraint` - where "str" is a unique string

A hybrid is defined in the module as a dictionary with the keys:

- `def` - the list of x,y coordinates given anti-clockwise
- `blc` - the x,y coordinates of the bottom left corner relative to the module
- `orientation` - set as top or bot
- `ASIC`
- `fid1` - the x,y coordinates of a fiducial to be measured
- `fid2` - the x,y coordinates of the other fiducial to be measured

An ASIC is defined in the hybrid as a dictionary with keys:

- `def` - the list of x,y coordinates given anti-clockwise
- `blc` - the x,y coordinates of the bottom left corner relative to the hybrid of the first ASIC in the array (left to right)
- `array` - the number of ASICs to be arrayed
- `spacing` - the spacing between ASICs in the array

The strip pads are dignified by the most extreme (left/right most) ones in the module as a dictionary with keys:

- `dim` - the width and height of a strip pad
- `top_left_pad` - the x,y coordinates of the pad relative to the ASIC
- `top_right_pad` - the x,y coordinates of the pad relative to the ASIC
- `bot_left_pad` - the x,y coordinates of the pad relative to the ASIC
- `bot_right_pad` - the x,y coordinates of the pad relative to the ASIC

Constraints can be defined such as the power_board as a dictionary in the module with keys:

- `def` - the list of x,y coordinates given anti-clockwise
- `blc` - the x,y coordinates of the bottom left corner relative to the module

All lists of coordinates (`def`) should be defined as tuples of (x,y) values in numpy arrays starting from the bottom left corner (`blc`) and going anti-clockwise. i.e.

```
"def":np.array([ (0.0,0.0), (0.0,17.0), (20.5,17.0), (22.0,15.5),
(97.5,15.5), (97.5,0.0) ])
```

A circle segment can be defined by adding and `("R",1000.0)` in sequence between the two enclosing points where the second number is the radius of curvature in mm. i.e.

```
"def":np.array([ (0.0,0.0), (0.0,17.0), (20.5,17.0), (22.0,15.5),
(97.5,15.5), ("R",1000.0), (97.5,0.0) ])
```

The bottom left corner (`blc`), `spacing`, `dim` and top/bot_left/right_pad should be inputted as a tuple. i.e.

```
"blc" : (1.89,28.235)
```

The fiducials `fid1` and `fid2` should also be inputted as a tuple in a numpy array. i.e.

```
np.array([1.0,1.0])
```

All numbers should be inputted as floats and are in mm.

An example can be found for the short strip module in `modules/ss.py` although it should be noted that some values in this file are placeholders and should be changed/checked. This includes: `fid1`, `fid2`, `strip_pad[dim]`, `strip_pad[top_left_pad]`, `strip_pad[top_right_pad]`, `strip_pad[bot_left_pad]`, `strip_pad[bot_right_pad]`, `bond_angle`.

Measurements are to be fed in with a csv file in the measurements directory, with the `input_file_name` path changed in `main.py`. It should have the structure: `"str"_hybrid,x1,y1,fid1,x2,y2,fid2` with no empty lines, an example is given in `measurements/test_measurements.csv`.

It's output is a JSON file in the measurements directory that will be uploaded to the database. It contains the information: technician-name, measurements:number, pass, max bonding angle, min bonding angle, collisions, collision points. An example can be seen `measurements/test_output.json`.

A drawing function has been included to draw the module however this is very slow and is only recommended for debugging purposes. It can be uncommented at the bottom of the main function.

```
st.draw(temp_coord)
```

Appendix B

Analysis: $t\bar{t}H$ All Hadronic

This appendix shows cut flows for the cuts shown in Tables 8.1 to 8.4 in Section 8.1. The alternative representation shown here highlights the impact of each successive cut.

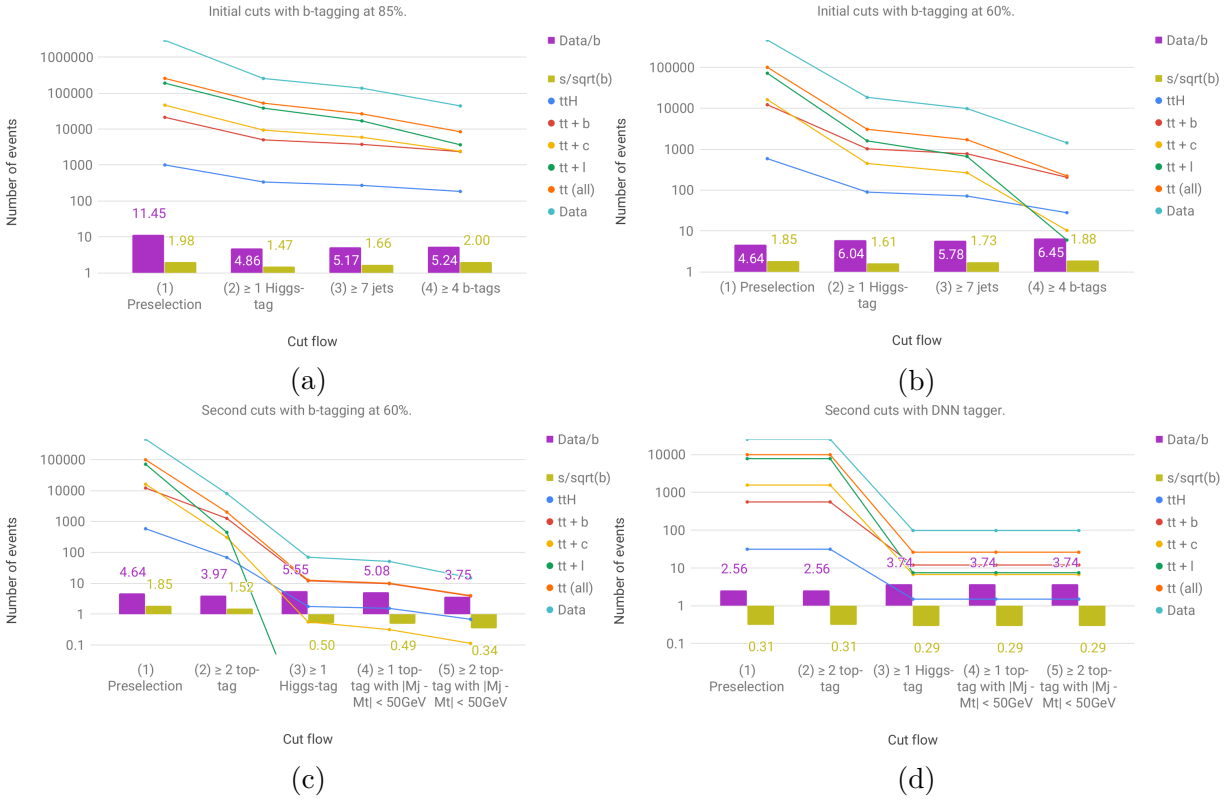


Figure B.1: Cut flows for the all hadronic feasibility study showing the $s\sqrt{b}$ and $data/b$ metrics.

Appendix C

Lepton + Jets Full Run-2 Analysis

This Appendix contains supporting material for the Lepton + Jets Full Run-2 Analysis in Section 8.2.

C.1 Analysis Regions

Figure C.1 details the successive cuts the define the signal and control regions for the Lepton + Jets analysis. Each signal region is split into the STXS bins which are defined in terms of reconstructed Higgs p_T using the given regions reconstruction algorithm.

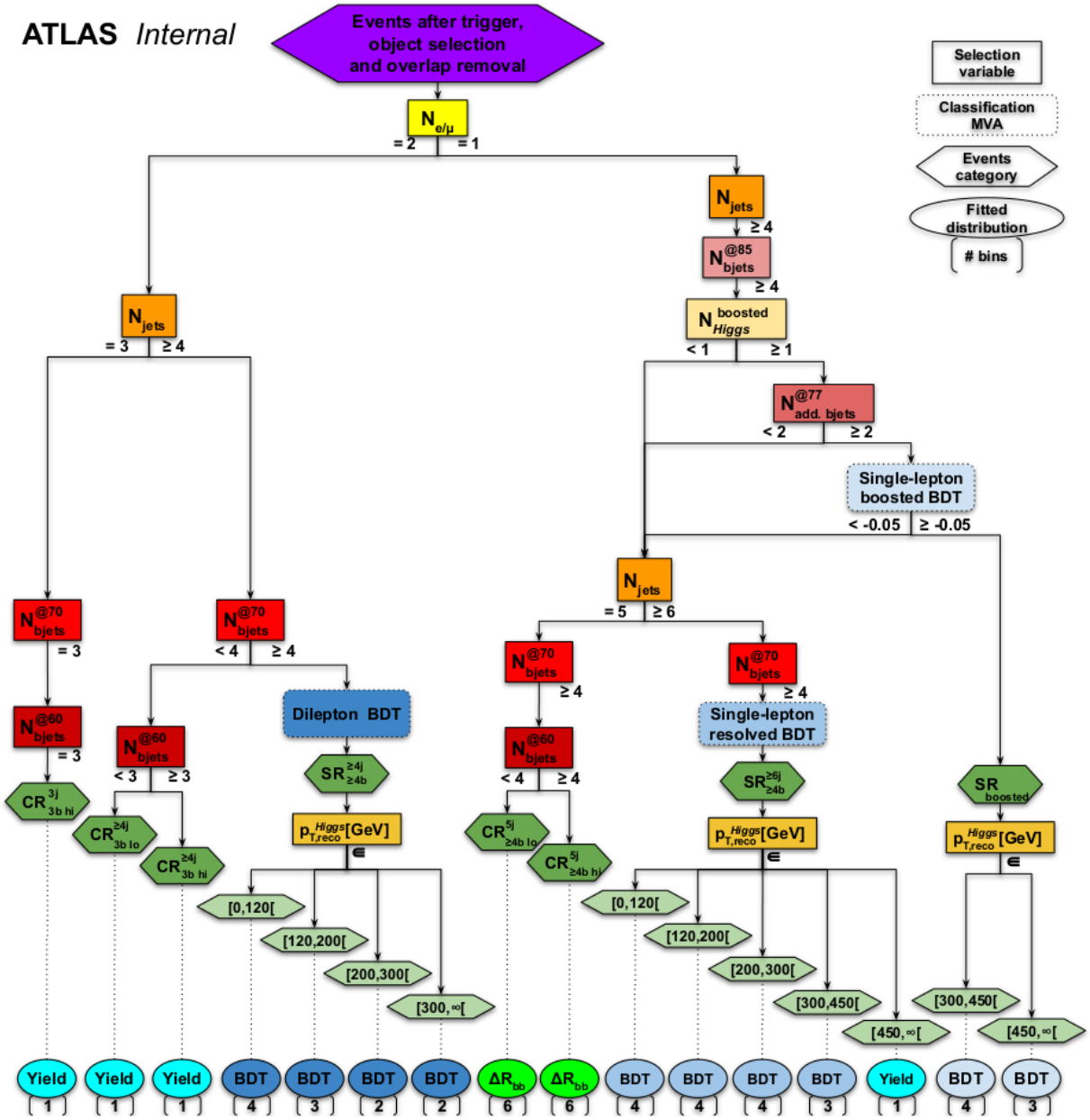


Figure C.1: Flow chart of the event selection for the Lepton + Jets Full Run-2 Analysis signal and control regions.

C.2 MC samples

Table C.1 shows the nominal and alternate MC samples used in the Lepton + Jets Full Run-2 Analysis in Section 8.2.

Process	ME generator	ME PDF	PS	Normalisation
Higgs boson				
$t\bar{t}H$	Powheg Box v2	NNPDF3.0$_{NLO}$	Pythia 8.230	NLO+NLO (EW) [37]
	Powheg Box v2	NNPDF3.0$_{NLO}$	Herwig 7.04	NLO+NLO (EW) [37]
	MadGraph5_@aMCNLO 2.6.0	NNPDF3.0$_{NLO}$	Pythia 8.230	NLO+NLO (EW) [37]
$tHjb$	MadGraph5_@aMCNLO 2.6.2	NNPDF3.0$_{NLO}$ nf4	Pythia 8.230	-
tWH	MadGraph5_@aMCNLO 2.6.2 [DR]	NNPDF3.0$_{NLO}$	Pythia 8.235	-
$t\bar{t}$ + jets and single-top				
$t\bar{t}$	Powheg Box v2	NNPDF3.0$_{NLO}$	Pythia 8.230	NNLO+NNLL [111–117]
	Powheg Box v2	NNPDF3.0$_{NLO}$	Herwig 7.04	NNLO+NNLL [111–117]
	MadGraph5_@aMCNLO 2.6.0	NNPDF3.0$_{NLO}$	Pythia 8.230	NNLO+NNLL [111–117]
$t\bar{t}+b\bar{b}$	Powheg Box Res	NNPDF3.0$_{NLO}$ nf4	Pythia 8.230	-
	Sherpa 2.2.1	NNPDF3.0$_{NNLO}$ nf4	Sherpa	-
tW	Powheg Box v2 [DR]	NNPDF3.0$_{NLO}$	Pythia 8.230	NLO+NNLL [118, 119]
	Powheg Box v2 [DS]	NNPDF3.0$_{NLO}$	Pythia 8.230	NLO+NNLL [118, 119]
	Powheg Box v2 [DR]	NNPDF3.0$_{NLO}$	Herwig 7.04	NLO+NNLL [118, 119]
	MadGraph5_@aMCNLO 2.6.2 [DR]	CT10$_{NLO}$	Pythia 8.230	NLO+NNLL [118, 119]
t-channel	Powheg Box v2	NNPDF3.0$_{NLO}$ nf4	Pythia 8.230	NLO [120, 121]
	Powheg Box v2	NNPDF3.0$_{NLO}$ nf4	Herwig 7.04	NLO [120, 121]
	MadGraph5_@aMCNLO 2.6.2	NNPDF3.0$_{NLO}$ nf4	Pythia 8.230	NLO [120, 121]
s-channel	Powheg Box v2	NNPDF3.0$_{NLO}$	Pythia 8.230	NLO [120, 121]
	Powheg Box v2	NNPDF3.0$_{NLO}$	Herwig 7.04	NLO [120, 121]
	MadGraph5_@aMCNLO 2.6.2	NNPDF3.0$_{NLO}$	Pythia 8.230	NLO [120, 121]
Other				
W+jets	Sherpa 2.2.1 (NLO [2j], LO [4j])	NNPDF3.0$_{NNLO}$	Sherpa	NNLO [122]
Z+jets	Sherpa 2.2.1 (NLO [2j], LO [4j])	NNPDF3.0$_{NNLO}$	Sherpa	NNLO [122]
VV (had.)	Sherpa 2.2.1	NNPDF3.0$_{NNLO}$	Sherpa	-
VV (lep.)	Sherpa 2.2.2	NNPDF3.0$_{NNLO}$	Sherpa	-
VV (lep.) + jj	Sherpa 2.2.2 (LO [EW])	NNPDF3.0$_{NNLO}$	Sherpa	-
$t\bar{t}W$	MadGraph5_@aMCNLO 2.3.3	NNPDF3.0$_{NLO}$	Pythia 8.210	NLO+NLO (EW) [37]
	Sherpa 2.0.0 (LO [2j])	NNPDF3.0$_{NNLO}$	Sherpa	NLO+NLO (EW) [37]
$t\bar{t}ll$	MadGraph5_@aMCNLO 2.3.3	NNPDF3.0$_{NLO}$	Pythia 8.210	NLO+NLO (EW) [37]
	Sherpa 2.0.0 (LO [1j])	NNPDF3.0$_{NNLO}$	Sherpa	NLO+NLO (EW) [37]
$t\bar{t}Z(qq, vv)$	MadGraph5_@aMCNLO 2.3.3	NNPDF3.0$_{NLO}$	Pythia 8.210	NLO+NLO (EW) [37]
	Sherpa 2.0.0 (LO [2j])	NNPDF3.0$_{NNLO}$	Sherpa	NLO+NLO (EW) [37]
$t\bar{t}t\bar{t}$	MadGraph5_@aMCNLO 2.3.3	NNPDF3.1$_{NLO}$	Pythia 8.230	NLO+NLO (EW) [37]
tZq	MadGraph5_@aMCNLO 2.3.3 (LO)	CTEQ6L1	Pythia 8.212	-
tWZ	MadGraph5_@aMCNLO 2.3.3 (DR)	NNPDF3.0$_{NLO}$	Pythia 8.230	-

Table C.1: MC events used in the Lepton + Jets analysis Full Run-2 Analysis [1]. The first row for each process shows the nominal sample. Additional rows contain the alternant samples used to evaluate the modelling. The overlap between $t\bar{t}$ and tW-like diagrams are dealt with either through a diagram removal scheme [DR] [123] or diagram subtraction scheme [DS] [124]. Unless otherwise stated the ME is calculated to NLO in QCD. The last column refers to the normalisation used and is calculated in order of QCD. For VV Sherpa samples (lep.) and (had.) refer to either both or one boson decaying leptonically.

C.3 DNN VR Pre-/post-fit plots

Figures C.2 to C.11 show all the pre- and post-fit plots of the boosted Higgs reconstruction DNNs input features. Overall it shows good modelling with the Data/MC ratio within uncertainties. Please refer to Section 8.2.2 for a more detailed discussion of the worst (but still good) modelled variables.

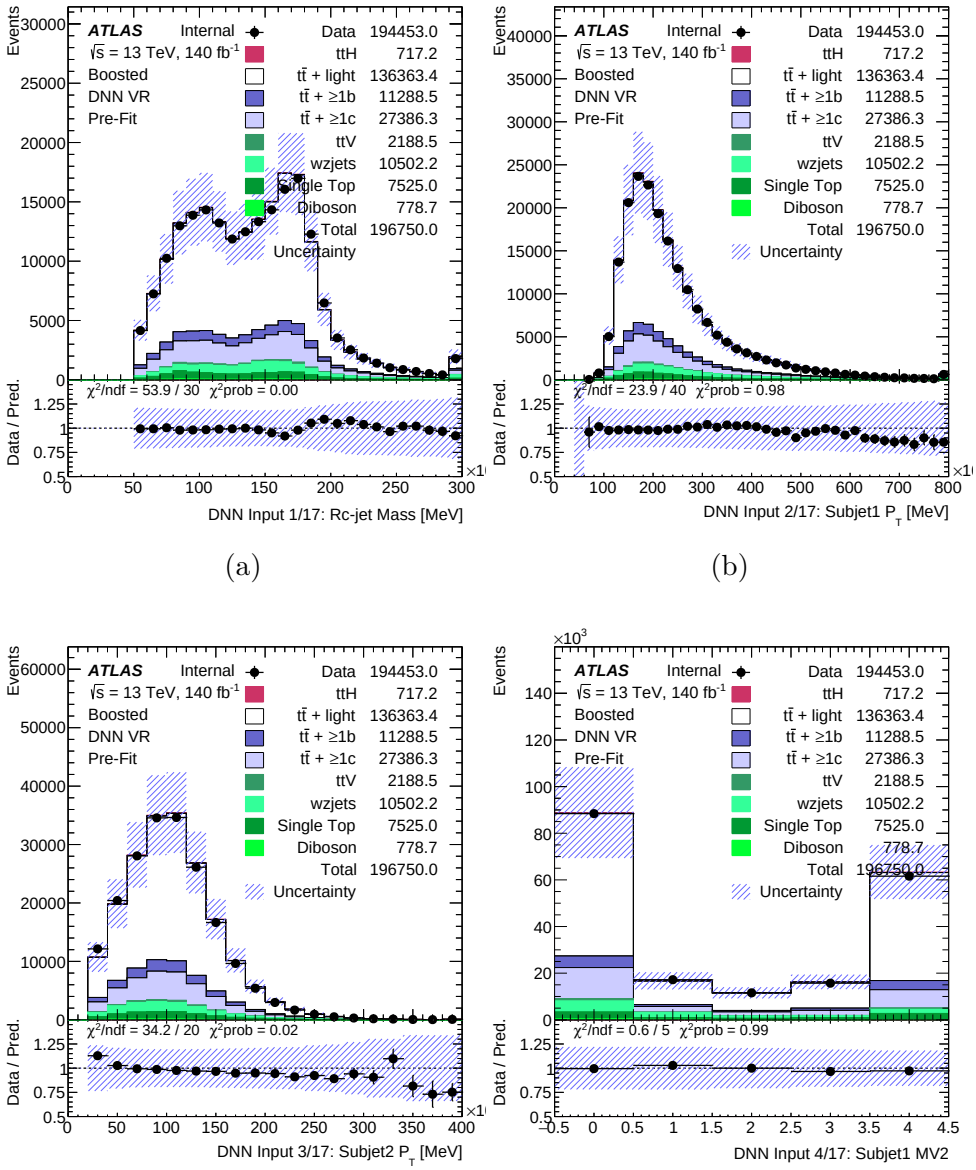


Figure C.2: Pre-fit Data/MC comparisons for DNN inputs variables: reclustered jet mass (C.2a), subjet 1 p_T (C.2b), subjet 2 p_T (C.2c), subjet 1 b-tagging score (C.2d).

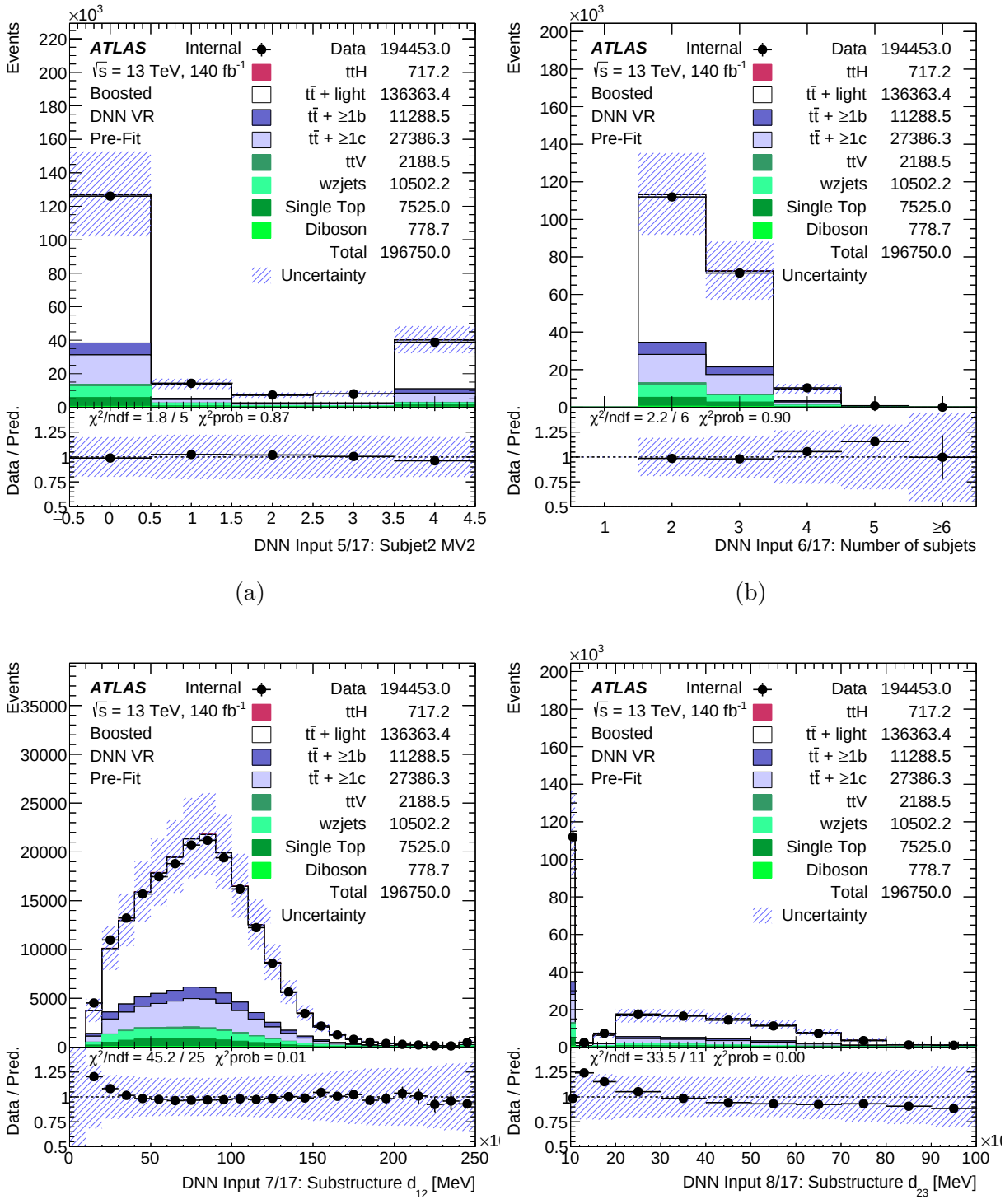
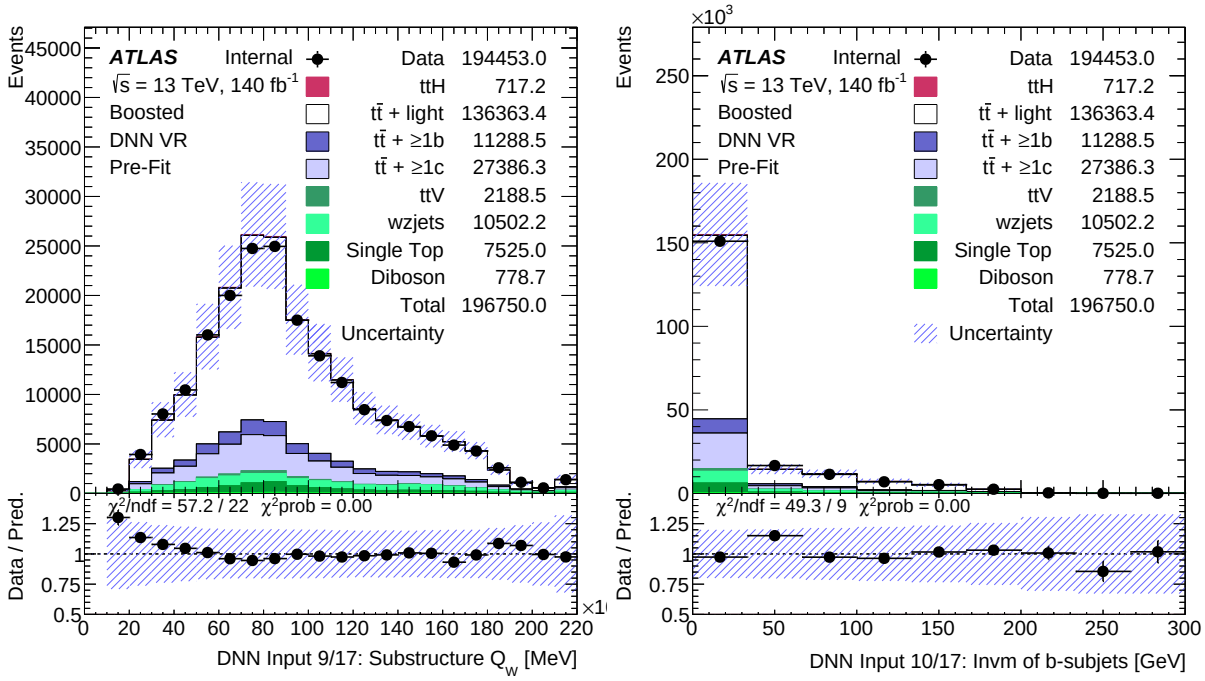


Figure C.3: Pre-fit Data/MC comparisons for DNN inputs variables: subjet 2 b-tagging score (C.3a), number of subjets (C.3b), substructure variable d_{12} (C.3c), substructure variable d_{23} (C.3d).



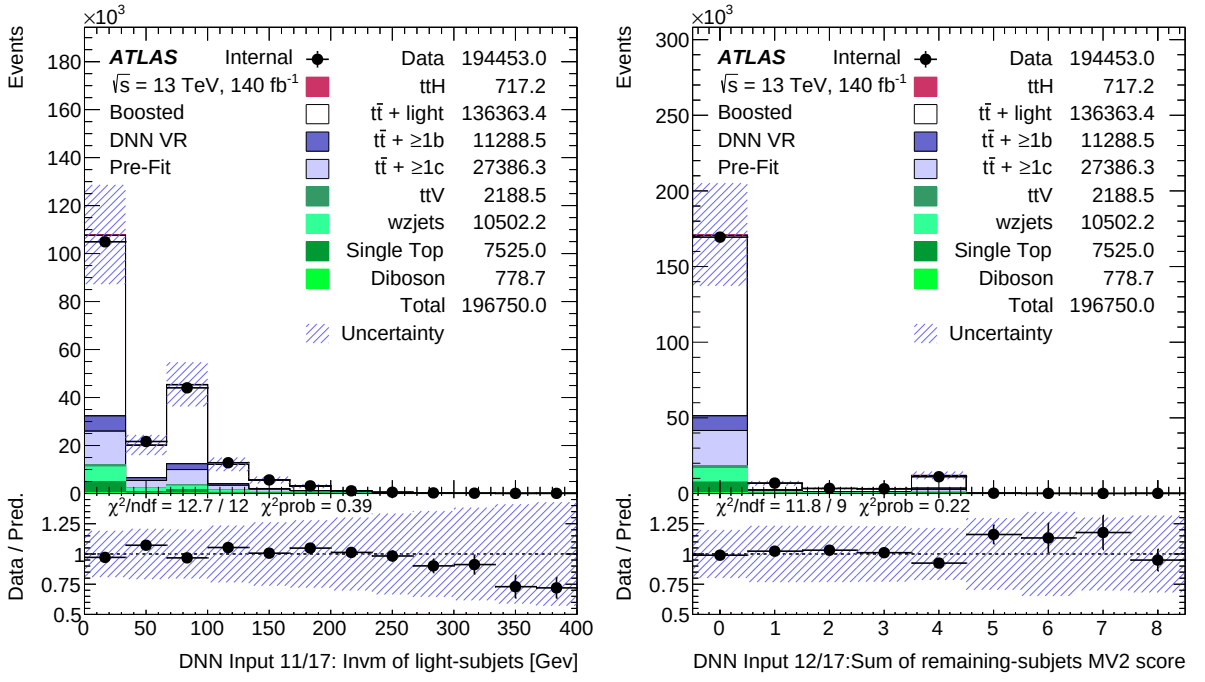


Figure C.4: Pre-fit Data/MC comparisons for DNN inputs variables: substructure variable Q_W (C.4a), invariant mass of b-tagged subjets (C.4b), invariant mass of light jets (C.4c), sum of remaining subjets b-tagging score (C.4d).

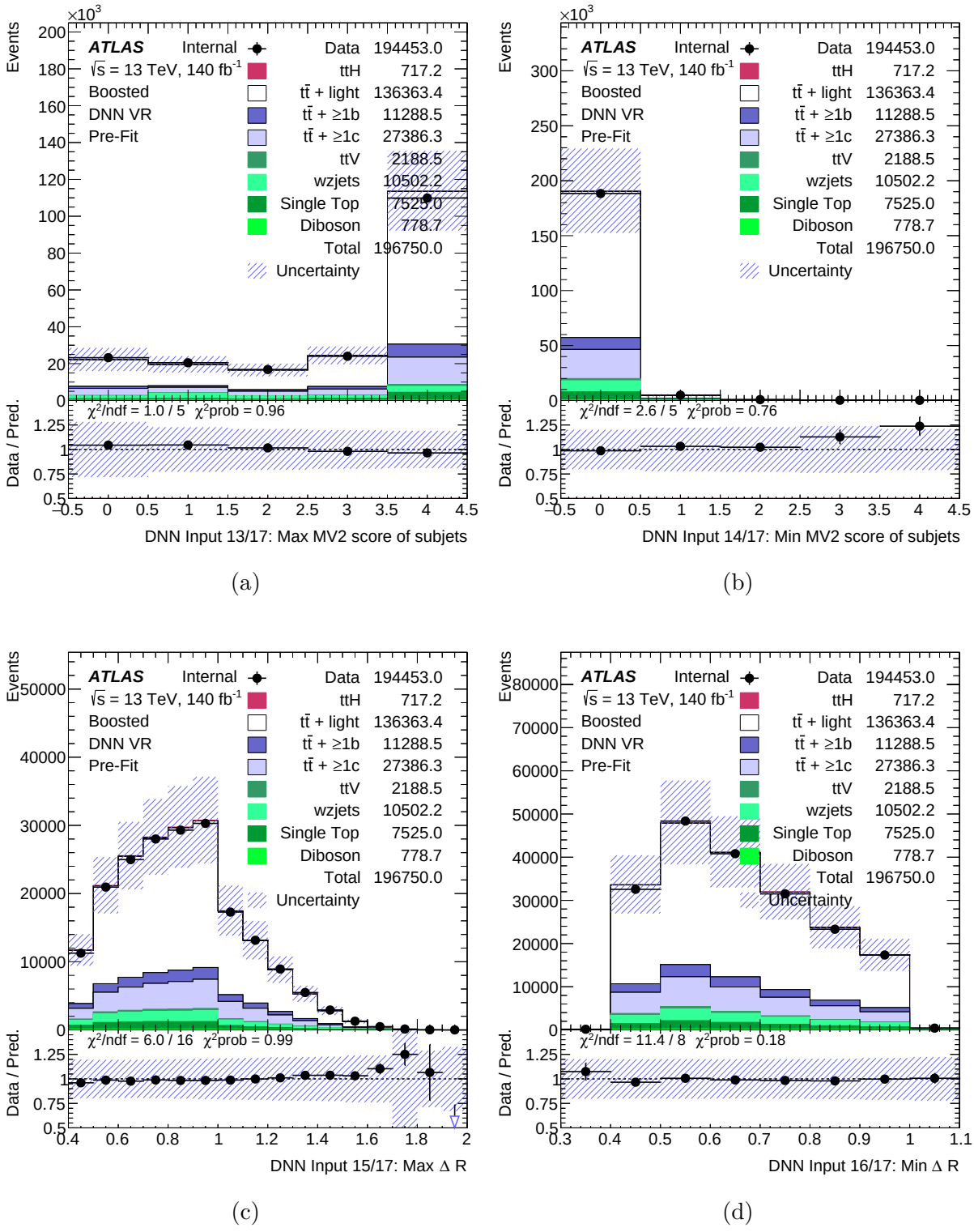


Figure C.5: Pre-fit Data/MC comparisons for DNN inputs variables: max b-tagging score of subjets (C.5a), min b-tagging score of subjets (C.5b), max ΔR of subjets (C.5c), min ΔR of subjets (C.5d).

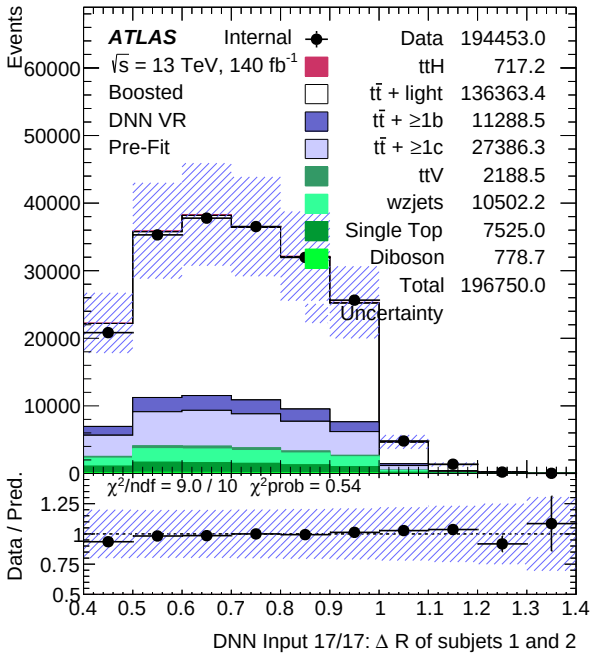
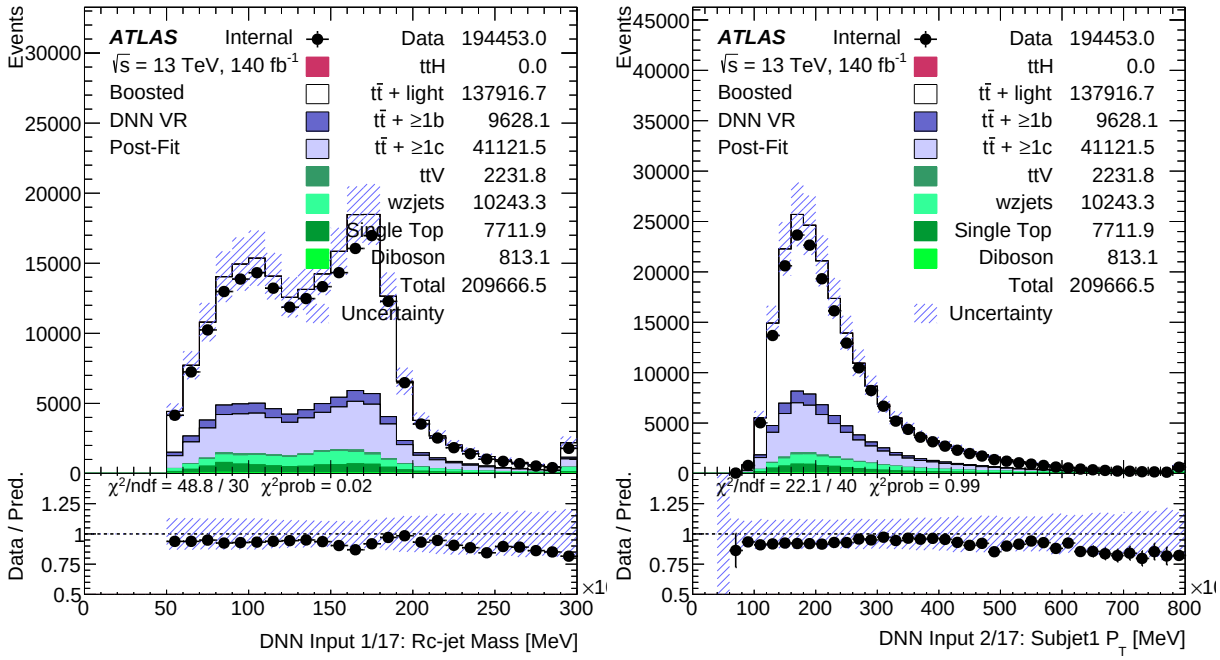
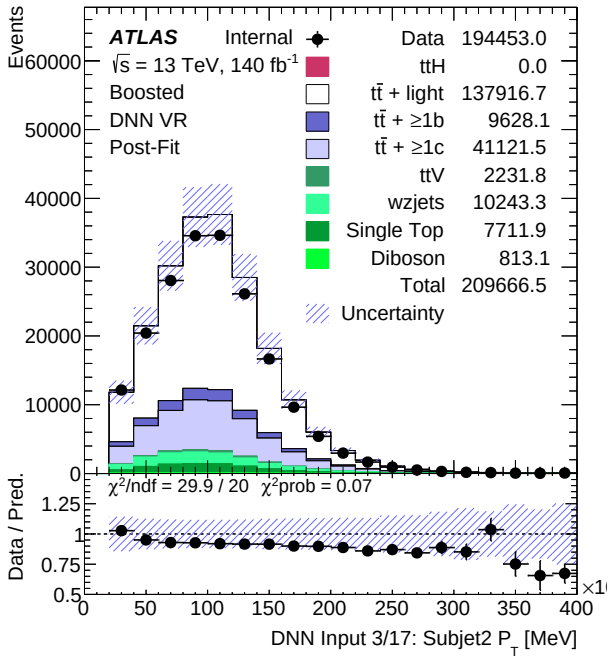


Figure C.6: Pre-fit ΔR of subjets 1 and 2.

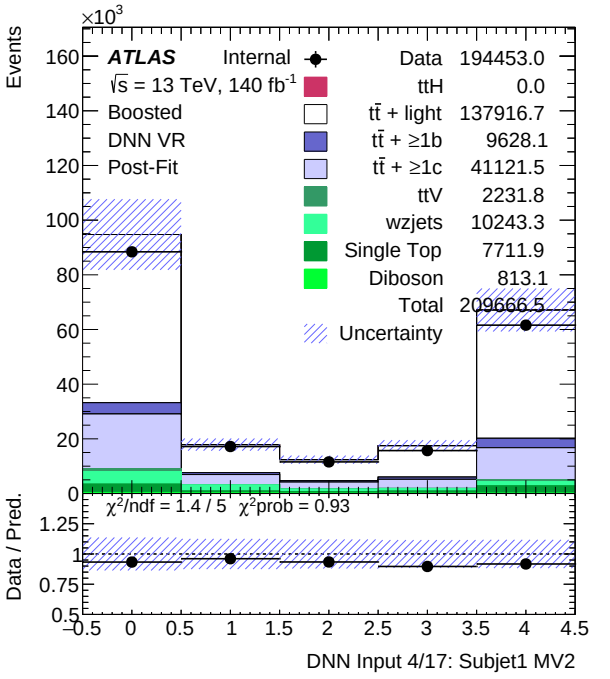


(a)

(b)

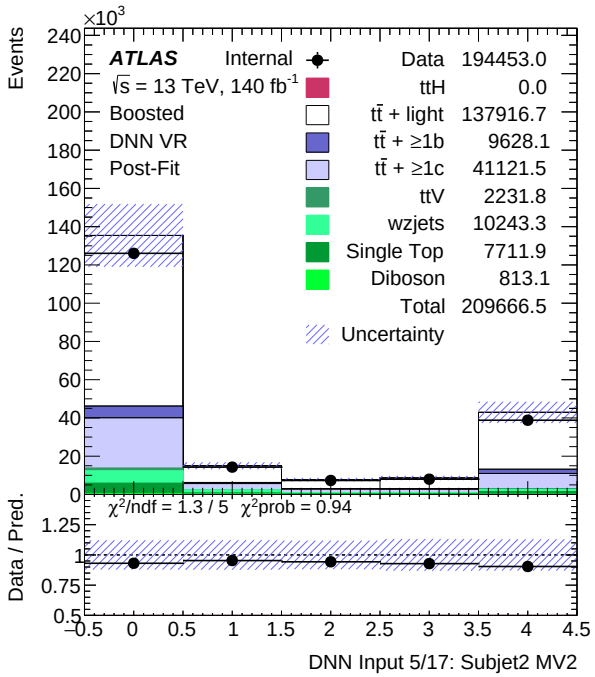


(c)

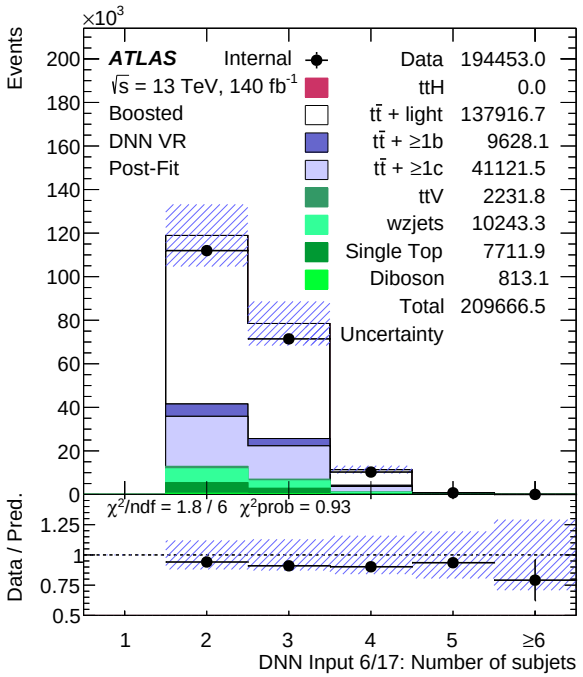


(d)

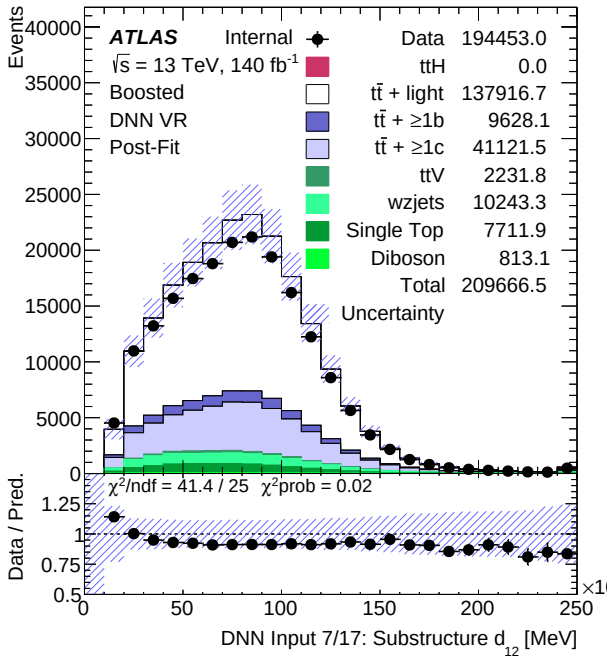
Figure C.7: Post-fit Data/MC comparisons for DNN inputs variables: reclustered jet mass (C.7a), subjet 1 p_T (C.7b), subjet 2 p_T (C.7c), subjet 1 b-tagging score (C.7d).



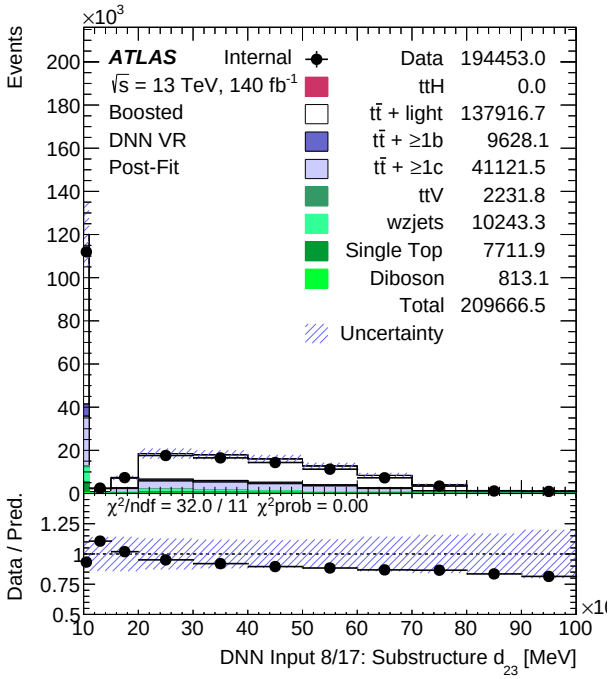
(a)



(b)



(c)



(d) Substructure variable d_{23}.

Figure C.8: Post-fit Data/MC comparisons for DNN inputs variables: subjet 2 b-tagging score (C.8a), number of subjets (C.8b), substructure variable d_{12} (C.8c), substructure variable d_{23} (C.8d).

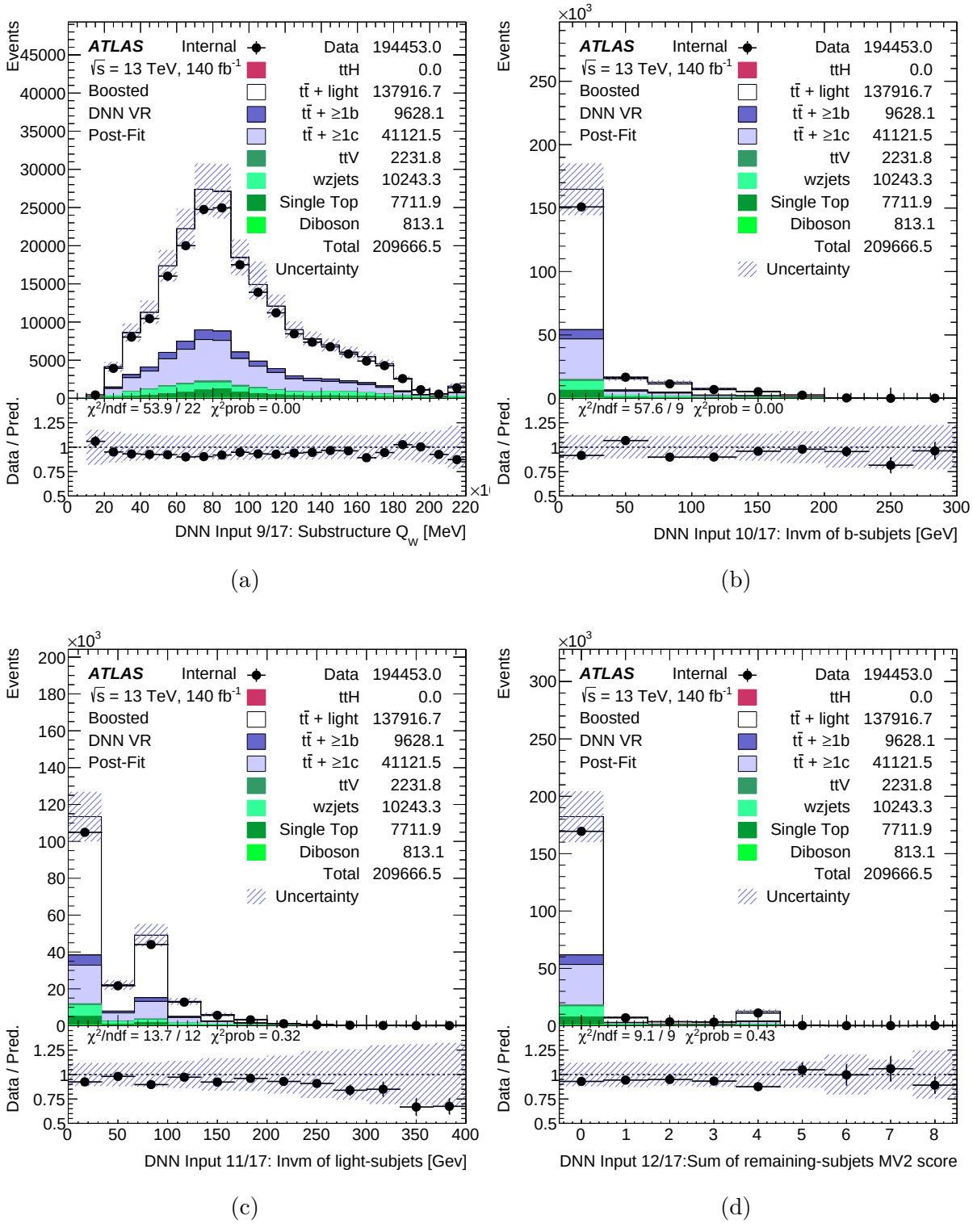
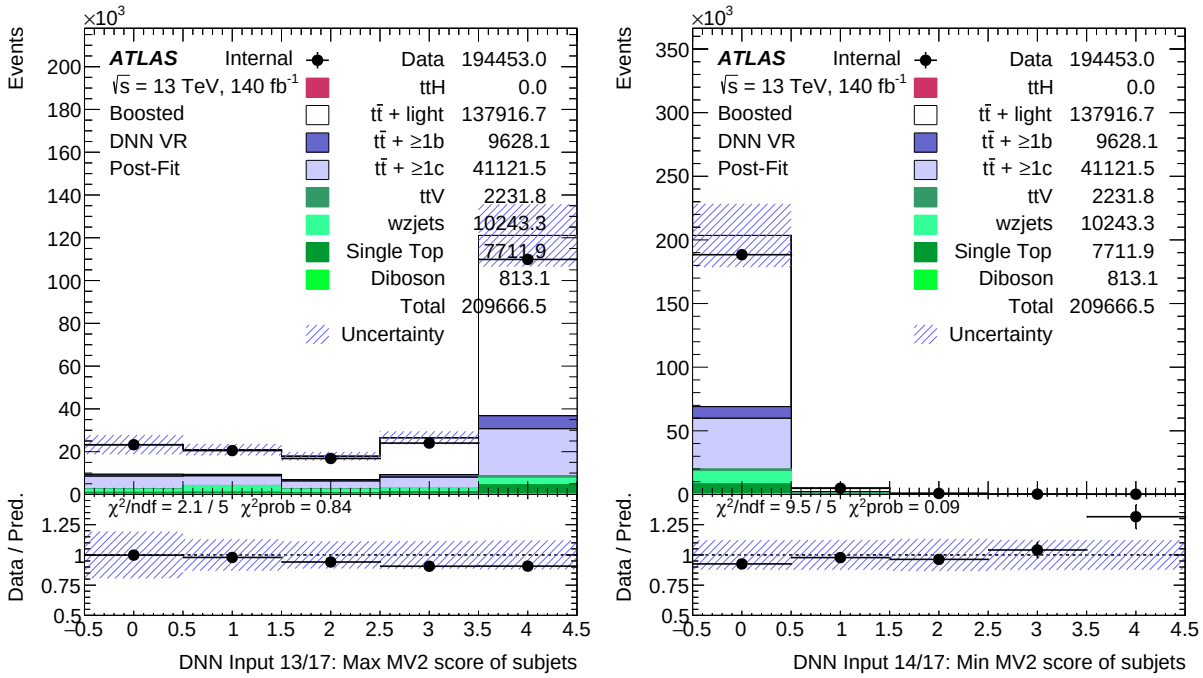
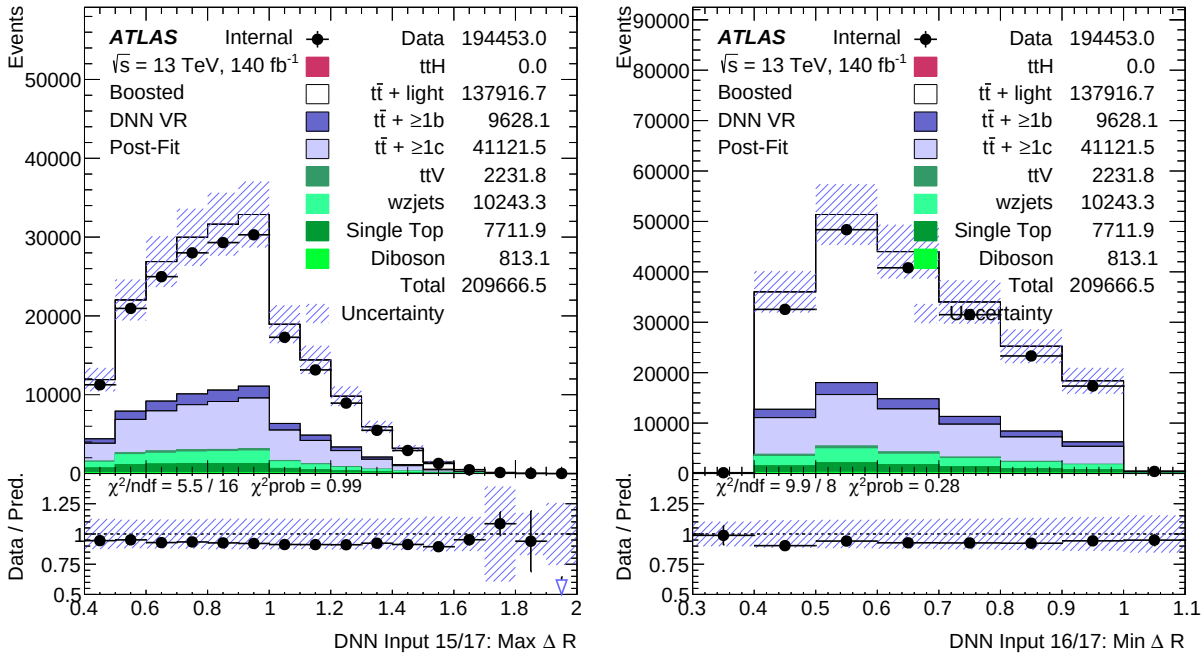


Figure C.9: Post-fit Data/MC comparisons for DNN inputs variables: substructure variable Q_W (C.9a), invariant mass of b-tagged subjets (C.9b), invariant mass of light jets (C.9c), sum of remaining subjets b-tagging score (C.9d).



(a) (b)



(c) (d)

Figure C.10: Post-fit Data/MC comparisons for DNN inputs variables: max b-tagging score of subjets (C.10a), min b-tagging score of subjets (C.10b), max ΔR of subjets (C.10c), min ΔR of subjets (C.10d).

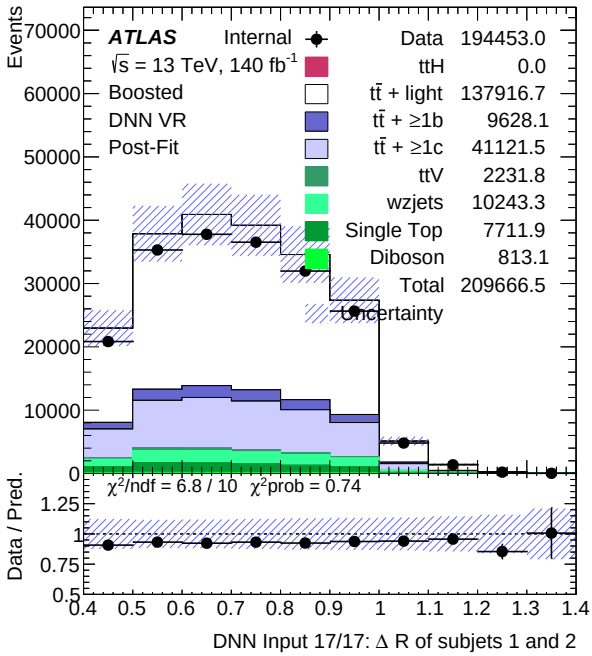
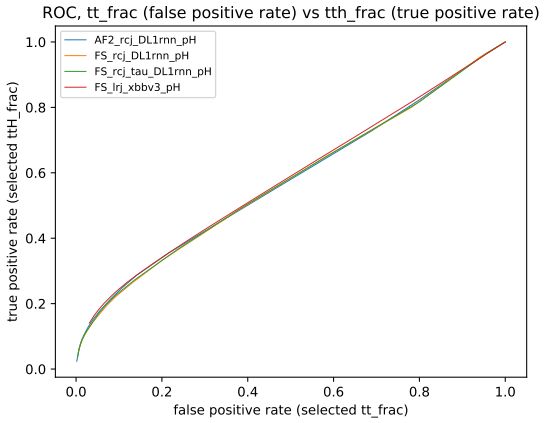


Figure C.11: Post-fit ΔR of subjets 1 and 2.

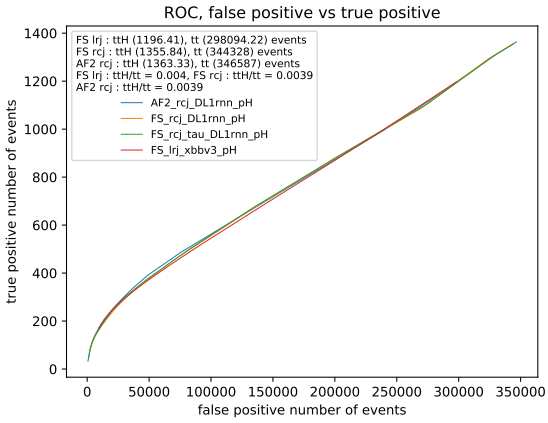
Appendix D

Lepton + Jets Legacy Analysis

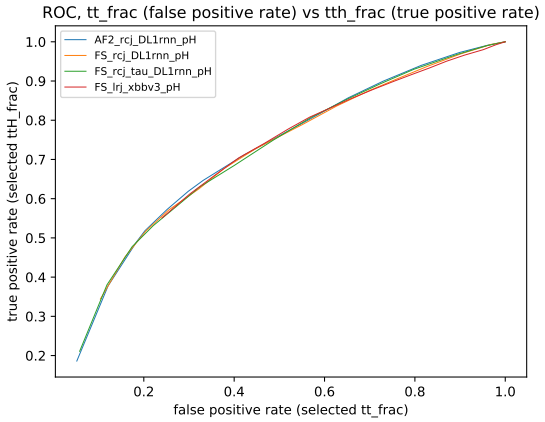
This Appendix contains the classic ROC curves in both relative and absolute number of events. They show the DNNs performances in both the Loose and Boosted regions as defined in section 8.3.



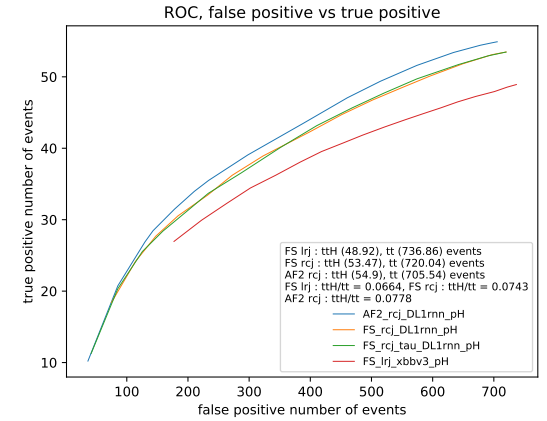
(a) TPR (relative) vs FPR on an event basis.



(b) TPR (absolute) vs FPR on an event basis.



(c) TPR (relative) vs FPR on an event basis.



(d) TPR (absolute) vs FPR on an event basis.

Figure D.1: ROC curves comparing DNN variations and the Xbb tagger in the Loose and Boosted regions.

Bibliography

[1] G. Aad *et al.*, "Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *JHEP*, vol. 06, p. 097, 2022.

[2] Wikipedia, "Standard model." `https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`.

[3] E. Noether, "Invariante variationsprobleme," *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, vol. 1918, pp. 235–257, 1918.

[4] S. L. Glashow, "The renormalizability of vector meson interactions," *Nuclear Physics*, vol. 10, pp. 107–117, 1959.

[5] A. Salam and J. C. Ward, "Weak and electromagnetic interactions," *Nuovo Cim.*, vol. 11, pp. 568–577, 1959.

[6] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[7] A. Salam and J. Ward, "Electromagnetic and weak interactions," *Physics Letters*, vol. 13, no. 2, pp. 168–171, 1964.

[8] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.

[9] J. Ellis, "Higgs Physics," in *2013 European School of High-Energy Physics*, pp. 117–168, 2015.

[10] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B*, vol. 716, pp. 1–29, 2012.

[11] S. Chatrchyan *et al.*, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys. Lett. B*, vol. 716, pp. 30–61, 2012.

[12] R. L. Workman and Others, "Review of Particle Physics," *PTEP*, vol. 2022, p. 083C01, 2022.

[13] M. Tanabashi *et al.*, "Review of Particle Physics," *Phys. Rev.*, vol. D98, no. 3, p. 030001, 2018.

[14] N. Cabibbo, "Unitary symmetry and leptonic decays," *Phys. Rev. Lett.*, vol. 10, pp. 531–533, Jun 1963.

[15] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction," *Progress of Theoretical Physics*, vol. 49, pp. 652–657, 02 1973.

[16] F. Abe *et al.*, "Observation of top quark production in $\bar{p}p$ collisions," *Phys. Rev. Lett.*, vol. 74, pp. 2626–2631, 1995.

[17] S. Abachi *et al.*, "Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV," *Phys. Rev. Lett.*, vol. 74, pp. 2422–2426, 1995.

[18] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge," *Zh. Eksp. Teor. Fiz.*, vol. 34, p. 247, 1957.

[19] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the Unified Model of Elementary Particles," *Progress of Theoretical Physics*, vol. 28, pp. 870–880, 11 1962.

[20] M. Aaboud *et al.*, "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector," *Phys. Lett.*, vol. B784, pp. 173–191, 2018.

[21] A. M. Sirunyan *et al.*, "Observation of $\mathrm{t\bar{t}H}$ production," *Phys. Rev. Lett.*, vol. 120, no. 23, p. 231801, 2018.

[22] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, "Higgs production in association with a single top quark at the lhc," *The European Physical Journal C*, vol. 75, no. 6, pp. 1–18, 2015.

[23] M. Aaboud *et al.*, "Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Rev. D*, vol. 97, no. 7, p. 072016, 2018.

[24] A. M. Sirunyan *et al.*, "Search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel with leptonic $\mathrm{t\bar{t}}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 03, p. 026, 2019.

[25] "Measurement of $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel in $41.5\,\mathrm{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$," 2019.

[26] A. collaboration, "Luminosity data." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[27] CERN, "Run-3 announcement." https://home.cern/news/news/physics/lhc-run-3-physics-record-energy-starts-tomorrow.

[28] G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, p. S08003, 2008.

[29] J. Pequenao and P. Schaffner, "How ATLAS detects particles: diagram of particle paths in the detector." 2013.

[30] "ATLAS Insertable B-Layer Technical Design Report Addendum," 2012.

[31] "Technical Design Report for the ATLAS Inner Tracker Strip Detector," 4 2017.

[32] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *JHEP*, vol. 02, p. 007, 2009.

[33] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and collider physics*. Cambridge university press, 2003.

[34] Z. Collaboration, "the zeus detector , technical proposal," tech. rep., DESY-HERAZEUS-1, März, 1986.

[35] H. Collaboration *et al.*, "Technical proposal for the h1 detector," *DESY, Hamburg*, p. 2, 1986.

[36] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Höche, H. Hoeth, F. Krauss, L. Lönnblad, E. Nurse, P. Richardson, S. Schumann, M. H. Seymour, T. Sjöstrand, P. Skands, and B. Webber, "General-purpose event generators for lhc physics," *Physics Reports*, vol. 504, no. 5, pp. 145–233, 2011.

[37] D. de Florian *et al.*, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector," vol. 2/2017, 10 2016.

[38] T. Sjöstrand, "Jet fragmentation of multiparton configurations in a string framework," *Nuclear Physics B*, vol. 248, no. 2, pp. 469–502, 1984.

[39] M. Bahr *et al.*, "Herwig++ Physics and Manual," *Eur. Phys. J. C*, vol. 58, pp. 639–707, 2008.

[40] S. Agostinelli *et al.*, "GEANT4–a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003.

[41] W. Lukas, "Fast Simulation for ATLAS: Atlfast-II and ISF," *J. Phys. Conf. Ser.*, vol. 396, p. 022031, 2012.

[42] T. Yamanaka, "The ATLAS calorimeter simulation FastCaloSim," *J. Phys. Conf. Ser.*, vol. 331, p. 032053, 2011.

[43] E. Gross and O. Vitells, "Trial factors for the look elsewhere effect in high energy physics," *The European Physical Journal C*, vol. 70, no. 1-2, pp. 525–530, 2010.

[44] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, 1938.

[45] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the American Mathematical society*, vol. 54, no. 3, pp. 426–482, 1943.

[46] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C*, vol. 71, no. 2, p. 1554, 2011.

[47] G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals," *Physical Review D*, vol. 57, no. 7, p. 3873, 1998.

[48] Y. Li, "Deep reinforcement learning: An overview," *arXiv preprint arXiv:1701.07274*, 2017.

[49] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[50] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[51] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[52] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[53] Wikipedia, "Ensemble boosting." `https://upload.wikimedia.org/wikipedia/commons/thumb/b/b5/Ensemble_Boosting.svg/1024px-Ensemble_Boosting.svg.png`.

[54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[55] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 315–323, JMLR Workshop and Conference Proceedings, 2011.

[56] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256, JMLR Workshop and Conference Proceedings, 2010.

[57] M. A. Nielsen, *Neural networks and deep learning*, vol. 25. Determination press San Francisco, CA, USA, 2015.

[58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[59] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural networks: Tricks of the trade*, pp. 437–478, Springer, 2012.

[60] M. Li, T. Zhang, Y. Chen, and A. J. Smola, "Efficient mini-batch training for stochastic optimization," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 661–670, 2014.

[61] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[62] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, *et al.*, "Wide & deep learning for recommender systems," in *Proceedings of the 1st workshop on deep learning for recommender systems*, pp. 7–10, 2016.

[63] F. C. et al., "Keras." `https://keras.io/api/`.

[64] M. A. et al., "Tensorflow." `https://www.tensorflow.org/api_docs`.

[65] D. H. G. et al., "Lightweight trained neural network (lwtnn)." `https://github.com/lwtnn/lwtnn`.

[66] G. D. Krause A., "Advanced topics in machine learning.." `http://courses.cms.caltech.edu/cs253/`, 2010.

[67] L. Rosasco, E. De Vito, A. Caponnetto, M. Piana, and A. Verri, "Are loss functions all the same?," *Neural computation*, vol. 16, no. 5, pp. 1063–1076, 2004.

[68] Wikipedia, "Cross-validation." `https://en.wikipedia.org/wiki/Cross-validation_(statistics)#/media/File:K-fold_cross_validation_EN.jpg`.

[69] G. Aad *et al.*, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1," *Eur. Phys. J. C*, vol. 77, p. 490, 2017.

[70] A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing," *J. ACM*, vol. 13, p. 471494, oct 1966.

[71] R. Frühwirth, "Application of kalman filtering to track and vertex fitting," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2, pp. 444–450, 1987.

[72] M. Aaboud *et al.*, "Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2," *Eur. Phys. J. C*, vol. 77, no. 10, p. 673, 2017.

[73] M. Aaboud *et al.*, "Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC," *Eur. Phys. J. C*, vol. 77, no. 5, p. 332, 2017.

[74] "Secondary vertex finding for jet flavour identification with the ATLAS detector," 6 2017.

[75] J. Illingworth and J. Kittler, "A survey of the hough transform," *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 1, pp. 87–116, 1988.

[76] G. Aad *et al.*, "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV," *Eur. Phys. J. C*, vol. 76, no. 5, p. 292, 2016.

[77] M. Aaboud *et al.*, "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 79, no. 8, p. 639, 2019.

[78] G. Aad *et al.*, "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data," *JINST*, vol. 14, no. 12, p. P12006, 2019.

[79] G. Aad *et al.*, "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 81, no. 7, p. 578, 2021.

[80] J. E. Huth, N. Wainer, K. Meier, N. Hadley, F. Aversa, M. Greco, P. Chiappetta, J. P. Guillet, and S. Ellis, "Toward a standardization of jet definitions,"

[81] G. P. Salam, "Towards Jetography," *Eur. Phys. J.*, vol. C67, pp. 637–686, 2010.

[82] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms," *JHEP*, vol. 08, p. 001, 1997.

[83] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[84] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur. Phys. J.*, vol. C72, p. 1896, 2012.

[85] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder," *Phys. Lett. B*, vol. 641, pp. 57–61, 2006.

[86] M. Aaboud *et al.*, "Jet reconstruction and performance using particle flow with the ATLAS Detector," *Eur. Phys. J. C*, vol. 77, no. 7, p. 466, 2017.

[87] G. Aad *et al.*, "Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Eur. Phys. J. C*, vol. 81, no. 8, p. 689, 2021.

[88] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Phys. Lett. B*, vol. 659, pp. 119–126, 2008.

[89] A. collaboration *et al.*, "Optimisation of the atlas b-tagging performance for the 2016 lhc run," tech. rep., ATL-PHYS-PUB-2016-012, 2016.

[90] G. Aad *et al.*, "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 79, no. 11, p. 970, 2019.

[91] "Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run," 7 2017.

[92] "Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC," 2018.

[93] G. Aad *et al.*, "Measurement of kT splitting scales in W->lv events at sqrt(s)=7 TeV with the ATLAS detector," *Eur. Phys. J. C*, vol. 73, no. 5, p. 2432, 2013.

[94] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness," *JHEP*, vol. 03, p. 015, 2011.

[95] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy Correlation Functions for Jet Substructure," *JHEP*, vol. 06, p. 108, 2013.

[96] P. Nason, "A new method for combining nlo qcd with shower monte carlo algorithms," *Journal of High Energy Physics*, vol. 2004, p. 040, dec 2004.

[97] S. Frixione, P. Nason, and C. Oleari, "Matching nlo qcd computations with parton shower simulations: the powheg method," *Journal of High Energy Physics*, vol. 2007, p. 070, nov 2007.

[98] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP*, vol. 06, p. 043, 2010.

[99] H. B. Hartanto, B. Jäger, L. Reina, and D. Wackeroth, "Higgs boson production in association with top quarks in the powheg box," *Phys. Rev. D*, vol. 91, p. 094003, May 2015.

[100] S. Frixione, G. Ridolfi, and P. Nason, "A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction," *Journal of High Energy Physics*, vol. 2007, p. 126, sep 2007.

[101] R. D. Ball *et al.*, "Parton distributions for the LHC Run II," *JHEP*, vol. 04, p. 040, 2015.

[102] G. Aad *et al.*, "Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," *JHEP*, vol. 05, p. 160, 2016.

[103] G. Aad *et al.*, "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data," *JINST*, vol. 14, no. 12, p. P12006, 2019.

[104] G. Aad *et al.*, "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C*, vol. 81, no. 7, p. 578, 2021.

[105] M. Aaboud *et al.*, "Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector," *Phys. Rev. D*, vol. 97, no. 7, p. 072003, 2018.

[106] A. H. et al., "Trexfitter." `https://trexfitter-docs.web.cern.ch/trexfitter-docs/`.

[107] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, "The roostats project." `https://arxiv.org/abs/1009.1003`, 2010.

[108] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, "The roostats project." `https://arxiv.org/abs/1009.1003`, 2010.

[109] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola, "Deep sets," 2017.

[110] S. Agostinelli *et al.*, "GEANT4–a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003.

[111] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, "Hadronic top-quark pair production with nnll threshold resummation," *Nuclear Physics B*, vol. 855, no. 3, pp. 695–741, 2012.

[112] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation," *Physics Letters B*, vol. 710, no. 4, pp. 612–622, 2012.

[113] P. Bärnreuther, M. Czakon, and A. Mitov, "Percent-level-precision physics at the tevatron: Next-to-next-to-leading order qcd corrections to $q\bar{q} \to t\bar{t}+x$," *Phys. Rev. Lett.*, vol. 109, p. 132001, Sep 2012.

[114] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels," *JHEP*, vol. 12, p. 054, 2012.

[115] M. Czakon and A. Mitov, "NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction," *JHEP*, vol. 01, p. 080, 2013.

[116] M. Czakon, P. Fiedler, and A. Mitov, "Total top-quark pair-production cross section at hadron colliders through $\mathcal{O}(\alpha_S^4)$," *Phys. Rev. Lett.*, vol. 110, p. 252004, Jun 2013.

[117] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders," *Comput. Phys. Commun.*, vol. 185, p. 2930, 2014.

[118] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-," *Phys. Rev. D*, vol. 82, p. 054018, 2010.

[119] N. Kidonakis, "Top Quark Production," in *Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons*, pp. 139–168, 2014.

[120] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann, "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR," *Comput. Phys. Commun.*, vol. 182, pp. 1034–1046, 2011.

[121] P. Kant, O. M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz, P. Rieck, and P. Uwer, "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions," *Comput. Phys. Commun.*, vol. 191, pp. 74–89, 2015.

[122] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, "High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO," *Phys. Rev. D*, vol. 69, p. 094008, 2004.

[123] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, "Single-top hadroproduction in association with a W boson," *JHEP*, vol. 07, p. 029, 2008.

[124] "Studies on top-quark Monte Carlo modelling for Top2016," 10 2016.