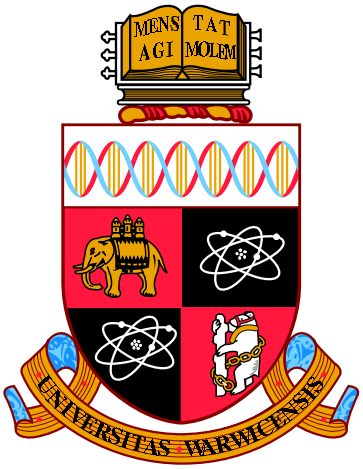



# Jets and their substructure at the ATLAS detector: New tools for new physics

by

**Joseph Stanford Ennis**

**Thesis**

Submitted to the University of Warwick

for the degree of

**Doctor of Philosophy**

**Department of Physics**

September 2017

THE UNIVERSITY OF WARWICK

# Contents

# List of Tables

# List of Figures

# Acknowledgments

Despite any evidence to the contrary found in the text of this thesis, I am not one to be overly grandiose or long-winded. Nevertheless, I must take the time to thank both the people without whom this thesis would truly not have been possible and those in whose absence it would have been substantially less fun.

First is surely my supervisors, Bill and Paul, under whose (mostly) wise tutelage I was given freedom to explore but enough guidance that I never lost my way, and who most likely solved more of my problems with an off-hand comment than I did with weeks of work. Secondly, the members of the Warwick ATLAS group and in particular, Tim and Graham, for their patient answering of my incessant questions at the beginning, and for their friendship since.

And to my family, I know I never quite convinced you it was a real job, but you were always there to support Rhiannon and me whenever things weren't at their best. I couldn't have come all this way without you.

Finally, and most importantly, to my wonderful wife Rhiannon. Thank you for all your love and patience over the last 4 years. Whatever the next adventure may be, I'm so happy we can explore it together.

# Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

It is a testament to the collaborative nature of experimental particle physics that no piece of work can be truly individual. The Large Hadron Collider and ATLAS experiment, on which I worked for the duration my studies, only function due to the hard work of thousands of people. Any work based on data collected at these, or similar, experiments is built upon the shoulders of many others. During my Ph.D., I made my own small contribution to the successful running of the ATLAS detector; through trigger service tasks and taking shifts overseeing the ATLAS trigger.

The work on multivariate jet calibration presented in §5 is entirely my own and that on jet triggers in §6 is primarily my own, The 2015 trigger efficiency study is my work but took place in the context of a wider group studying trigger performance. The substructure trigger study was initiated within a small group at the 2016 Jet Hackathon workshop where I contributed to all aspects of the analysis.

The $X \to VV \to qqqq$ analysis in §7, was the result of a continued collaboration between a group of physicists over the course of several years. My primary contributions were the implementation and maintenance of a significant portion of the analysis software, including event selection, calibration and histogramming tools, and the production of datasets. Additionally, I managed the data derivation strategy and analysis of the trigger behaviour, and I contributed a significant amount to

the study of $n_{trk}$ uncertainties. I was also responsible for the software used by other groups conducting similar analyses.

Except for those explicitly labelled, I produced all tables and figures included in this thesis.

# Abstract

This thesis is primarily concerned with two topics: the trigger performance and calibration of large-radius (large-R) jets at the ATLAS detector, and the presentation of a search for new high-mass particles that decay to vector bosons in the fully hadronic decay channel.

The performance of the ATLAS large-R jet trigger during 2015 was evaluated and found to be 100% efficient by 380 GeV for all large-R jets. The impact of jet substructure on this trigger was studied, and we discovered that a newly proposed set of variables, *anti-subjettiness variables*, could reduce trigger rates by approximately 30%, while still accepting 95% of large-R jets from heavy particles. Alongside this, a new calibration method for large-R jets is presented. A neural-network is used to perform a multivariate calibration for jet energy and mass using substructure variables. While there were no gains observed in the jet energy calibration, a 26% improvement in the jet mass resolution was found.

A search for $X \to VV \to qqqq$, using large-R jets tagged as boosted vector bosons, was conducted using 37.6 $fb^{-1}$ of data collected at $\sqrt{s} = 13$ TeV by ATLAS in 2015 and 2016. No significant deviation from the Standard Model was observed and limits were placed on a variety of generic signal models. The Heavy Vector Triplet model A (B) was excluded in the mass range $1.2(1.2) < m_V < 3.1(3.5)$ TeV, the RS graviton with $k/\overline{M}_{planck} = 1$ was excluded for $1.3 < m_G < 1.6$ TeV, and limits on $\sigma \times BR$ were placed on a generic scalar model.

# Chapter 1

# Introduction

''In the beginning the universe was created. This has made a lot of people very angry and been widely regarded as a bad move,'' or so claimed Douglas Adams in his seminal physics text *The Restaurant at the End of the Universe* [1]. Despite this, or, perhaps, in a concerted effort to prove him correct, scientists working on the Large Hadron Collider (or LHC) study particle interactions at energy levels similar to those from the first few seconds of the universe to try and gain ever greater understanding of its fundamental rules.

For the last few decades our theoretical knowledge of fundamental physics has been brought together in the so-called ''Standard Model'' of particle physics, whose predictions have driven much of the experimental work in this field. The Standard Model has predicted the existence of the Z boson, discovered in 1983 at LEP, the Large Electron Positron collider that was the original occupant of the tunnel now containing the LHC, [2], the top quark, observed in 1995 at the Tevatron proton-antiproton collider [3, 4] and most recently the Higgs boson, which was finally discovered in 2012 at the LHC, 48 years after it was proposed [5, 6]. The Standard Model has not just correctly predicted the existence of new particles; it has also provided accurate predictions for a huge variety of particle properties and interactions, at energy scales orders of magnitude apart.

However, while the Standard Model has proven astonishingly successful in many areas, it is obviously incomplete in others; for example, it does not yet include a mechanism for neutrino masses, and the theory breaks down with the addition of gravity. There are also parts of its structure that seem arbitrary, and many theorists believe they could be explained by the existence of some deeper, underlying theory.

In practice, there is no shortage of ideas on how to address these issues. Theories including supersymmetry [7], composite particles [8], leptoquarks [9], and more have been proposed to address the Standard Model's shortcomings; the challenge now lies not in theorising a way to fix the Standard Model, but in ascertaining which (if any) of an infinite spectrum of theoretical possibilities is correct.

It is possible that new physics may first be detected in subtle ways, minor deviations between the Standard Model predictions and measurements made at particle accelerators such as the LHC.[1] However, perhaps the most direct way would be to observe one of the many new particles predicted by ''Beyond the Standard Model'' (BSM) theories. Often, these are predicted to be much heavier than those that exist in the Standard Model and would decay almost instantaneously upon being created in a particle collision, almost undetectable amongst the noise. Thankfully, by using advanced detectors such as ATLAS, we may observe their remnants and attempt to reconstruct such decays, seeking the telltale signatures of a new particle amongst the noise.

The LHC provides a unique laboratory in which to search for new physics. When running, it accelerates bunches of protons to 7 TeV (or lead ions to 1.38 TeV) in two contra-rotating beams and continuously collides millions of protons each second at four main interaction points situated around the ring. At each interaction point there is a detector gathering collision data using highly complex calorimetry, tracking, triggering and data acquisition systems. ATLAS and CMS are both general purpose detectors with a near-hermetic cylindrical design, LHCb is a forward detector designed to study b-quark decays, and ALICE is optimised for the lead ion runs and studying the quark-gluon plasma produced in such collisions.

This thesis is primarily concerned with an analysis conducted on data collected by the ATLAS detector in 2015 and 2016 [11]. Rather than searching for particles associated with a specific theory we instead performed a search for a more generic class of particles that could be produced by a variety of models. The analysis was optimised to find particles that interact with the weak force and that decay, at least some of the time, to a pair of electroweak bosons; focusing on the final state that is produced by such a decay when both bosons then decay hadronically (analyses using other decay channels were also conducted [12–14]). No observation of a new particle was made, but the data collected allowed us to place constraints on the existence of such particles.

[1] In fact, the first hints of something new may have already been detected in violation of lepton universality through decays of $B \to K^* \mu\mu$ [10], though it is not yet statistically significant.

The hadronic decay of the bosons in this analysis presents an interesting challenge when reconstructing a collision. The strong force confines the quarks, and when one is produced by an interaction, many further quark-anti-quark pairs are created alongside it. These create further pairs (and so on) until the remaining particles finally form stable colour-neutral hadrons. Therefore, any quark produced through a high-energy process at the LHC is observed not as a single particle but as a spray of highly collimated hadrons called ''jets''. In a hadron collider, such as the LHC, the $pp \to jets$ cross-section is an order of magnitude higher than any other process and would easily drown out any possible signal. To make this final state even more challenging, the two jets produced by high-energy boson decays are themselves highly-collimated and often overlap.

To identify boson-like jets amongst the noise, large-radius jets and their substructure are used. Such techniques are becoming common in exotic searches at ATLAS and CMS [15–18] but are still undergoing rapid development. The secondary focus of this thesis is on performance studies with large-radius jets.

Since the LHC generates far more data than could possibly be stored, there is a ''trigger'' system responsible for analysing events in (close-to) real-time and selecting those containing interesting physics for storage. Which events are stored is a crucial choice for physics analyses; if the correct data is not collected then no analysis can be performed at all! As part of a larger study of the ATLAS trigger system [19], the performance of the ATLAS large-radius jet trigger was evaluated in data, using a bootstrap method from lower energy triggers, and its efficiency threshold was found.

However, increases in the luminosity delivered by the LHC have continually forced the trigger threshold to increase to prevent a significant rise in the trigger acceptance rate. The total rate of data-taking is fundamentally limited by the availability of output bandwidth and long-term storage, and could not be increased sufficiently to cope, requiring the trigger to throw away some data that would have previously been stored. We demonstrate that this can be mitigated by using an ''anti-QCD'' substructure-based tagger in the trigger, delivering a 30% reduction in trigger rate without significant impact on the collection of heavy particles such as the W or top quark.

Finally, this thesis will present an examination of the use of jet substructure when calibrating the large-radius jet mass and energy. This is not feasible with the methods currently used for ATLAS jet calibrations and required the development of a neural-network based calibration procedure. A proof-of-concept of this method is

presented alongside performance measurements. To select inputs from the large library of available substructure variables, two systematic methods for the selection of input variables were developed. It was found that the energy calibration does not benefit from this multivariate approach but when it was tested on the jet mass an improvement of 26% was obtained in the jet mass resolution.

The structure of this thesis is as follows. §2-4 are primarily a reference to the reader: §2 will describe the Large Hadron Collider, the CERN accelerator complex, and the ATLAS detector, including its trigger and data acquisition (DAQ) systems. §3 will give an overview of the Standard Model: its successes, problems, and the BSM theories that will be referenced in later chapters. §4 covers jets, in both theory and experiment, including generating hadrons in Monte-Carlo simulation, jet algorithms and a short introduction to jet substructure. §5-6 are primarily my own work and cover large-radius jet performance studies completed during my Ph.D., multivariate calibration studies, and ATLAS jet trigger performance respectively. Finally §7 presents the primary analysis on which I worked during my Ph.D., a search for heavy particles decaying to vector bosons in the fully hadronic channel. The chapter gives an overview of the entire analysis with a focus on my contributions.

# Chapter 2

# ATLAS and the LHC


Figure 2.1: "Look! Small rock come out of big rock! Small rock fundamental!" "...small rock is maybe statistical artifact." *Saturday Morning Breakfast Cereal: To The Collider!* [20].

There is an essential element that separates the scientific from the philosophical - the principle of empiricism. The success of a theory must be judged by observation, not by appeals to *a priori* reasoning or nebulous concepts such as elegance or beauty (though these can certainly be motivations in its development). In particle physics, this poses a unique challenge as we are concerned not with macroscopic behaviours or large ensembles of particles, but with the fundamental building blocks of nature and the laws governing their interactions at the smallest scales. Unsurprisingly, the

process of observation is somewhat tricky.

Directly probing a particle and its interactions requires our experiments to have a resolution of the same order as the size of the particle. As technology and experimental techniques have improved, we have been able to drill down into what we believed to be fundamental and discover new structure inside, often alongside new physical laws that explained previously unexplainable behaviour. The transition from a model of many fundamental atoms to atoms made of nucleons and electrons is one obvious example of this; the reduction of the ''particle zoo'' of hadrons to a set of quark bound-states another. In fact, thanks to de-Broglie, we can even quantify the limits of our observations and hence our knowledge. Given a particle of momentum $p$, the length-scale we can probe is set by $\lambda = \frac{h}{p}$.

To observe ever-smaller (and potentially more fundamental) particles, we require a corresponding increase in the energy of the particles with which we observe them. This places a fundamental limit on what can be measured using microscopes and light, set by the frequency range of the colour spectrum. To move beyond this requires higher-energy photons or observations using accelerated massive particles such as the electron or proton.[1]

This brings us, in a rather roundabout way, to the topic of this chapter, the proton-accelerating Large Hadron Collider (LHC) and the ATLAS detector.

## 2.1 The Large Hadron Collider

The LHC is a 27km circumference proton-proton/heavy ion collider at the CERN experimental complex. It was originally designed as a high-luminosity, low-energy partner to the planned Superconducting Supercollider in the US but following the cancellation of that project it has become the focal point of high-energy particle physics research [21]. The design goal for the LHC was a 14 TeV centre-of-mass-energy collider with a luminosity of $10^{34}\text{cm}^2\text{s}^{-1}$; while this luminosity target was achieved (and exceeded) in 2016, the LHC has not yet reached design energy, and is currently operating at 13 TeV.

[1] ...and hopefully one day soon the muon.

### 2.1.1 The Accelerator Complex

The LHC was designed to make use of the pre-existing complex of accelerators at CERN, shown in figure 2.2, and was built in the existing LEP tunnel to minimise costs. When in operation, protons are acquired by stripping the electrons from hydrogen atoms with a strong electric field once every 100$\mu$s. These protons are generated in bunches and for the protons this is the start of a long journey until they are finally injected into the LHC. After an initial acceleration by a high-intensity radio-frequency cavity (RFG2) the protons are accelerated to 50 MeV by the LINAC2 accelerator[2], then fed into the proton synchotron booster (PSB), built in 1972 to increase the maximum luminosity delivered by the proton synchotron, and the proton synchotron, (PS) where they are circulated until reaching 1.4 GeV and 25 GeV respectively. Once the protons have been accelerated by the PS, they are fed into the significantly larger super proton synchotron (SPS) and accelerated to 450 GeV, ready to be injected into the LHC.

Both the PS and SPS have a significant history prior to their use delivering protons to the LHC; both were flagship accelerators in their own right. The SPS was the site of the UA1 and UA2 experiments which discovered the W and Z bosons in 1983. The PSB was built after the PS to raise the injection energy of protons, allowing a 100-fold increase in luminosity within the PS.

Once they reach the LHC, a maximum of 2808 proton bunches, each containing $1.15 \times 10^1 1$ protons, are accelerated in contra-rotating beams until they reach the collision energy. of 6.5 TeV per-beam.

### 2.1.2 Circulating Beams

Accelerating, maintaining, and colliding the beams within the LHC is a complex process. The LHC beam ring is a single pipe containing two parallel beam lines that share cooling and magnetic systems. This structure can be seen in the cross-section from the dipole sections of the beam line in figure 2.3.

Once protons are injected into the main ring, acceleration is provided by 8 radio-frequency (RF) cavities laid along a straight section of the LHC beam. Each RF cavity provides 2 MeV of energy to a proton bunch on each pass through, for a total

[2]During the 2$^{\text{nd}}$ long shutdown period the LINAC2 accelerator will be replaced by the recently completed LINAC4 in preparation for the HL-LHC. LINAC4 will accelerate $H^-$ ions (rather than the $H^+$ currently used) to 160 MeV and will allow a doubling of intensity for the LHC feeder beam [22].

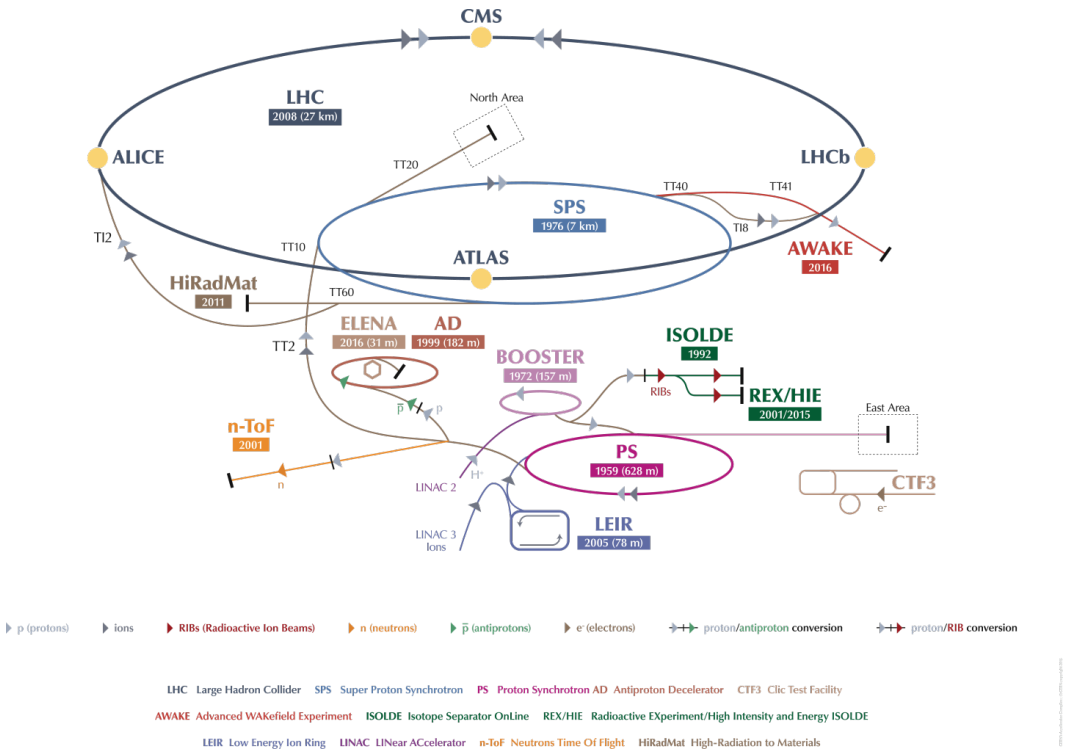



Figure 2.2: The accelerator complex at CERN. Protons for the LHC begin their life at the LINAC 2 accelerator and pass through a series of progressively more powerful accelerators until they reach their final destination. Reproduced from [23].

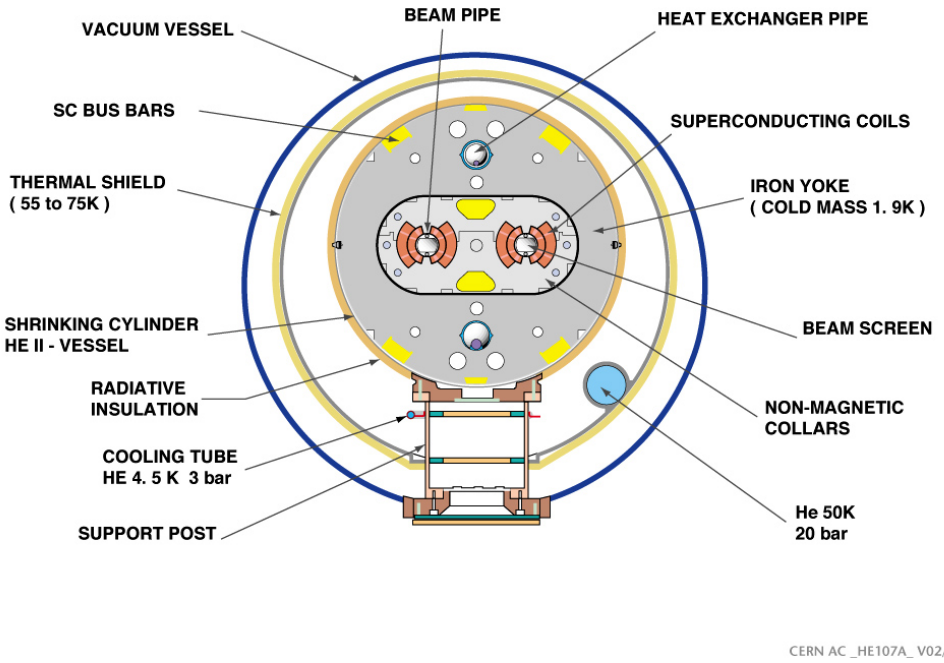



Figure 2.3: Cross-section of an LHC dipole section from [24].

of 16 MeV energy per cycle. The radio frequency used sets the bunch structure of the beams; to maintain acceleration, the phase of the electric field in the RF cavity must be the same each time a proton bunch passes through. This requires that the field frequency is an integer multiple of the revolution frequency of the bunches. In the case of the LHC, the RF frequency is 400MHz which is 35640 times the revolution frequency. This splits the LHC beam into 35640 possible ‘‘buckets’’ which may be filled by proton bunches; the LHC design specifies 2808 filled buckets, but 2015 and 2016 operation used a maximum of 2220.

The LHC magnet system consists of 1232 8.3T superconducting dipole magnets that bend the beams around the ring and 858 quadrupole magnets arranged in a FODO structure for focusing the beam. There are also higher order multi-pole magnets to correct for additional beam effects. These are NbTi superconducting magnets and must be cooled to 2K by the liquid helium cryogenics system. The strength of the dipole magnets was a significant parameter in the design of the LHC as for a fixed accelerator radius, $r$, (such as the existing LEP tunnel radius), the maximum possible energy is limited by the strength of the magnetic field, $B$,

$$E \propto Br. \tag{2.1}$$

Therefore the LHC's 14 TeV design energy is fixed by the 8.3T magnetic fields generated by the dipole magnets. Sadly, the LHC magnets have not yet achieved their design strength, and the collision energy is currently limited to 13 TeV (6.5 TeV per beam).

Magnets are also used for manipulating the beam within the ring. A dedicated ''beam dump'' rapidly removes the beam from the detector in the event of a fault or end-of-run. This is required to be extremely fast in the case of a fault as the beam could cause catastrophic damage if it intersected with a detector or other sensitive piece of equipment. In the event of a dump, the beam is redirected by 15 ''kicker'' magnets into a carbon/steel absorber. There are also eight sets of quadrupole magnets called the ''inner triplets'' are used to focus and defocus the beams around each interaction point, reducing the beta function and ''squeezing'' the beams to increase the collision rate. For ATLAS and CMS, the beta function is reduced from 11m to 0.4m at the interaction point.

There are eight beam cross-over points in the LHC ring where the beams could collide. Four of these, at Points 1,2,5 and 8 are the homes of the four primary LHC experiments, ATLAS, ALICE, CMS, and LHCb.

### 2.1.3 LHC Operation

In addition to collision energy, the LHC design goal was high luminosity. The event rate for a process is given by

$$R = \sigma \mathcal{L}, \tag{2.2}$$

where $\sigma$ is the process cross-section and $\mathcal{L}$ is the detector luminosity, so maximising $\mathcal{L}$ increases the probability of observing an interesting event and hence improves the sensitivity of many measurements made at the LHC. The relationship between many accelerator parameters and the (instantaneous) delivered luminosity is given by

$$\mathcal{L} = \frac{1}{4\pi} f_{rev} n_B N_B^2 \frac{\gamma}{\beta^* \epsilon_N} R \tag{2.3}$$

where $f_{rev}$ is the revolution frequency, $n_B$ is the number of bunches, $N_B$ is the number of particles in a bunch, $\beta^*$ is the beta function at the collision point, $\epsilon_N$ is the normalised emittance at the collision point and $R$ is a geometrical factor taking into account the crossing angle and further effects. $f_{rev}$, $R$ and $\gamma$ are fixed by the LHC design so the luminosity is manipulated at the LHC by changing the number of protons available for collision, $n_B N_B$, and the focusing of the beams, $\beta^* \epsilon_N$.

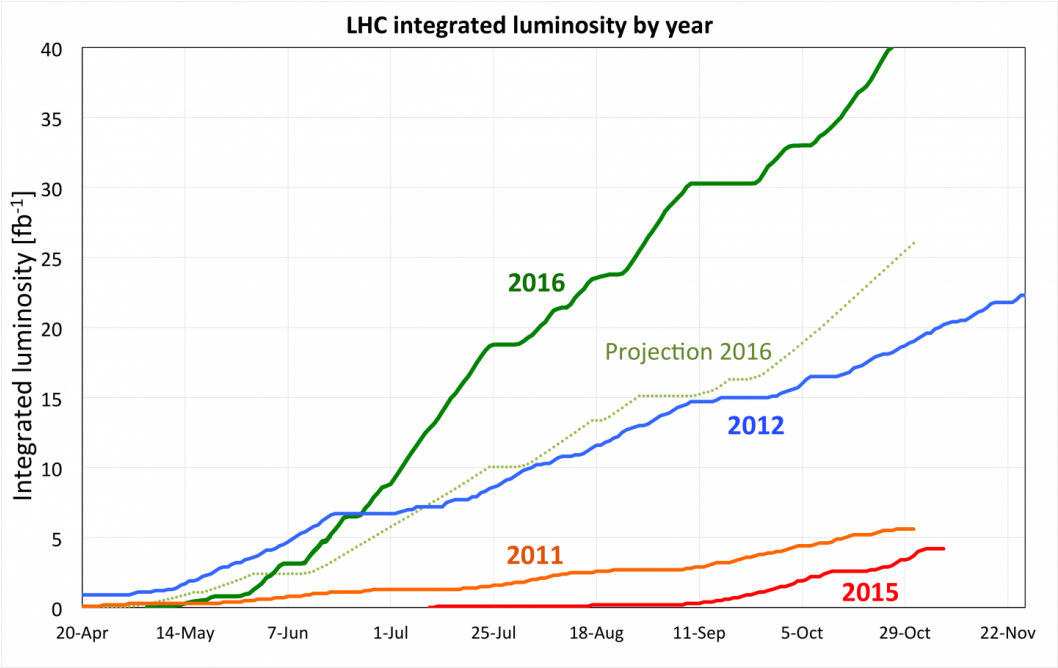



Figure 2.4: The integrated luminosity delivered per-year by the LHC for 2011-2016. Reproduced from [25].

Of course, (2.3) is only the instantaneous luminosity, as the beams circulate through the ring they will lose protons to both collisions and unwanted interactions within the beam pipe. The beam lifetime $\zeta$ is the expected time for the luminosity to fall to $1/e$ of its initial value and is around 10 hours in the LHC. The total luminosity delivered by the LHC is shown in figure 2.4.

## 2.2 The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) detector is a $4\pi$, general-purpose detector located at Point 1 on the LHC ring and the largest general-purpose particle detector ever constructed. When whole, the detector is 47m long, 25m wide and 25m high. It is divided into three sections, a large central barrel structure built parallel to the LHC beam pipe, and two fitted end-caps that provide coverage in the highly forward (and backward) regions.

The detector possesses an onion-like structure, similar to the CMS detector and previous general-purpose detectors, where each layer (or group of layers) was designed to fulfil a different purpose. The inner-most layers provide accurate tracking, the

electromagnetic and hadronic calorimeters both measure energy deposition but are tuned for the optimal measurement of electrons and photons, and hadronic particles respectively, and finally a dedicated muon spectrometer surrounds the interior layers. Magnetic fields are generated within the detector to allow charge identification and momentum measurement. A cryogenic system cools parts of the detector to required operational temperatures.

The intense luminosity delivered by the LHC causes significant challenges for the detector's design and operation, even beyond the physical requirement of radiation hardening. Reading out and storing the detector state for each of the 40 million collisions per second would be an insurmountable task. The information read-out from the detector during a typical collision is around 1.5Mb so the data collected during a single 8-hour run would require approximately 36Pb of storage space! To meet this challenge, the ATLAS detector has a complex data acquisition and trigger system. A complete description of the detector's design can be found in [26, 27].

### 2.2.1 Coordinate System

A standard coordinate system is used when discussing collisions within ATLAS. The z-axis is defined as the beam direction, and the x-y axes form a plane perpendicular to this with the x-axis pointing towards the centre of the LHC ring and the y-axis pointing directly "up" towards the surface. However, this $(x, y, z)$ coordinate system is not ideal for describing locations within ATLAS' cylindrical geometry. When analysing collisions we (typically) care only about where the particle was produced and its direction and location relative to the primary collision vertex. These can be specified by two angles, $\theta$, the polar angle from the beam axis, and $\phi$, the azimuthal angle around the beam axis. In order to make the coordinate system invariant under boosts along the z-axis (and hence the collision's location along the beam-line) the pseudorapidity $\eta = -\ln\tan\frac{\theta}{2}$ is used instead of $\theta$. $\eta$ is 0 when a particle is exactly transverse to the beam-line and becomes asymptotically infinite as it approaches the beam-line. In this pseudorapidity-azimuthal space, distance is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

### 2.2.2 Magnetic Fields

ATLAS contains four separate superconducting magnetic systems, a central solenoid surrounding the inner detector, an air-core toroid located within the muon spectrom-

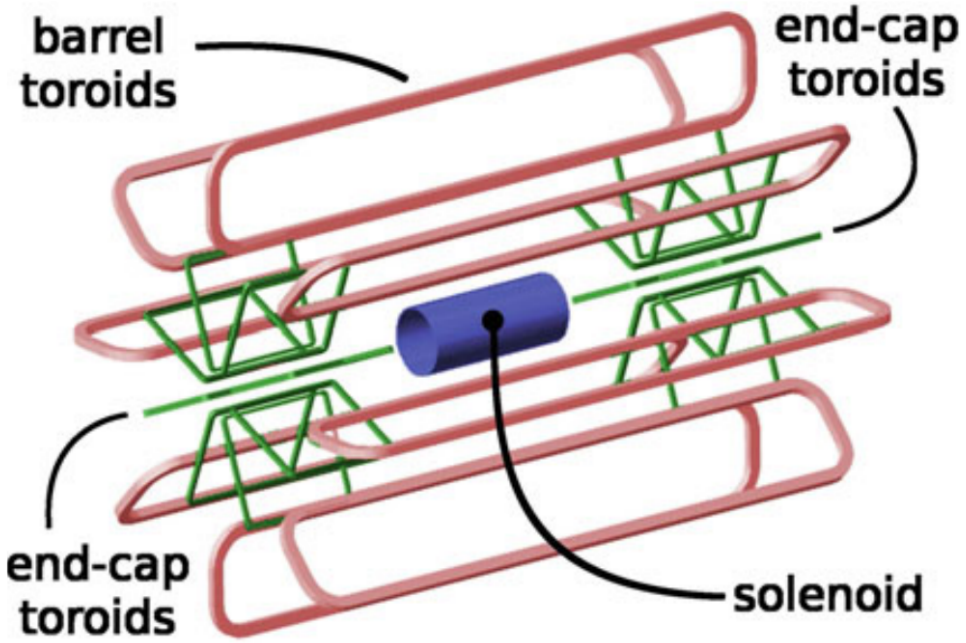



Figure 2.5: The ATLAS magnet systems. Figure from [28].

eter, and two further air-core toroids attached to the end-caps, which lie within the barrel as shown in figure 2.5. The combined magnet system is 26m long and 20m in diameter.

The central solenoid is a 5.3m long cylinder with an interior (exterior) diameter of 2.44m (2.64m) and a typical magnetic field strength of 2.0T. Its design was constrained by its location directly in front of the calorimeter system, which limited how thick it could be made. Each toroid has a radially symmetric layout of eight rings, and the end-cap toroids are rotationally offset from the barrel toroids to provide radial overlap and provide a magnetic field of approximately 2T inside the detector. The magnetic systems must be kept cool due to the superconducting coils used for both the solenoid and toroid systems, the magnets are cooled by the cryogenic system to an operation temperature of 4.5K.

### 2.2.3 Inner Detector

The inner detector is (predictably) the innermost segment of the ATLAS detector, located closest to the beamline and sitting within the 2T magnetic field provided by the central solenoid. It was designed to deliver high-precision tracking of charged

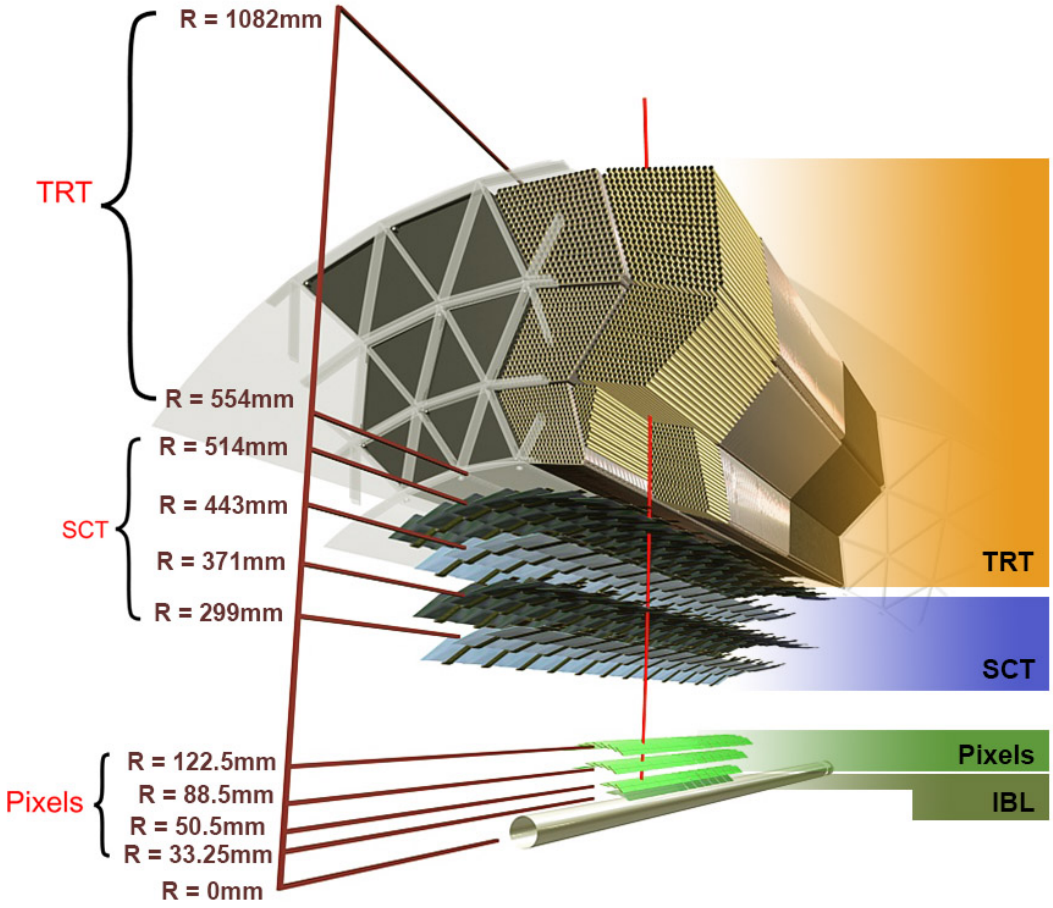
| Detector | Component | Length [m] | Radius [m] | $\lvert\eta\rvert$ Coverage |
|---|---|---|---|---|
| Inner Detector | Barrel | 3.5 | 1.2 | $\lvert\eta\rvert < 2.5$ |
| EM Calorimeter | Barrel | 2.8 | 1.4 - 2.0 | $\lvert\eta\rvert < 2.5$ |
| | End-cap | 0.3 | 2.1 | $1.4 < \lvert\eta\rvert < 3.2$ |
| Tile Calorimeter | Barrel | 5.8 | 2.3 - 4.3 | $\lvert\eta\rvert < 1.7$ |
| | Extended Barrel | 2.6 | 2.3 - 4.3 | $\lvert\eta\rvert < 1.7$ |
| Hadronic End-cap | - | 1.8 | 2.0 | $1.5 < \lvert\eta\rvert < 3.2$ |
| Forward Calorimeter | - | 1.3 | 0.5 | $3.1 < \lvert\eta\rvert < 4.9$ |

Table 2.1: A summary of the size and approximate $\eta$-coverage of the main detector subsystems.

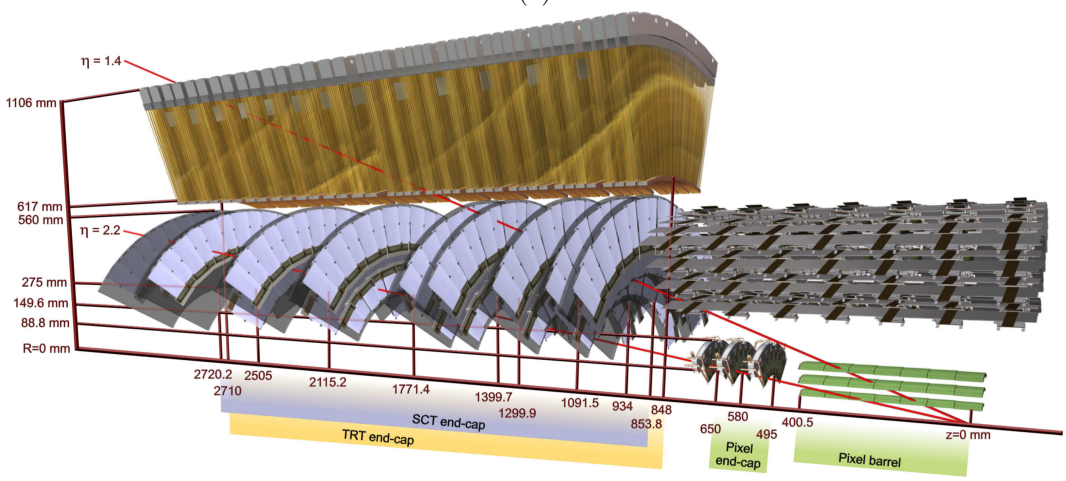
particles in a highly irradiated and noisy environment. As well as providing increased momentum resolution, precision tracking is a key requirement of operation at high-pileup as it aids discrimination between particles emitted from the primary interaction vertex and pile-up. The full inner detector provides high-precision tracking within $|\eta| < 2.5$ and is 6.2m long with a diameter of 2.1m; like the main detector, it is divided into barrel and end-cap segments to maximise coverage in pseudorapidity. The inner detector is split into three subsystems; moving outwards from the beamline they are the pixel detector, the silicon strip tracker (SCT), and the transition radiation tracker (TRT). During long-shutdown 1, a period from early 2013 to early 2015 when the LHC was shut down to perform significant upgrades to both collider and detectors, the inner detector was upgraded by the insertion of an additional pixel layer, called the insertable b-layer (IBL) [29], between the existing pixel layers and the beam pipe. This improved tracking performance and provides additional resilience to radiation damage.

The barrel pixel detector consists of 4 layers of semiconductor ‘‘pixels’’, sensors with a thickness of $250\mu m$ (200-$230\mu m$) and area of $50 \times 400\mu m^2$ ($50 \times 250\mu m^2$) in the original layers (IBL), positioned 3.3-12.2cm from the beam pipe. In the end cap are three discs with pixels similar to the outer barrel layers. There are a total of 92 million pixels, and the tracking resolution is $8 \times 40\mu m$ in the IBL and $15 \times 115\mu m$ in the outer layers.

Outside the pixel detector lies the SCT, which provides four layers of silicon microstrip sensors in the barrel and nine further layers in each endcap. The 4088 SCT strip detectors are laid out to provide close-to-hermetic coverage of the interaction point and deliver a resolution of $17 \times 580\mu m$. To operate at design capabilities in such an irradiated environment both SCT and pixel detector require active cooling




(a)




(b)

Figure 2.6: The layout of the inner detector barrel including the new IBL (a) [29] and end-cap (b) [27].

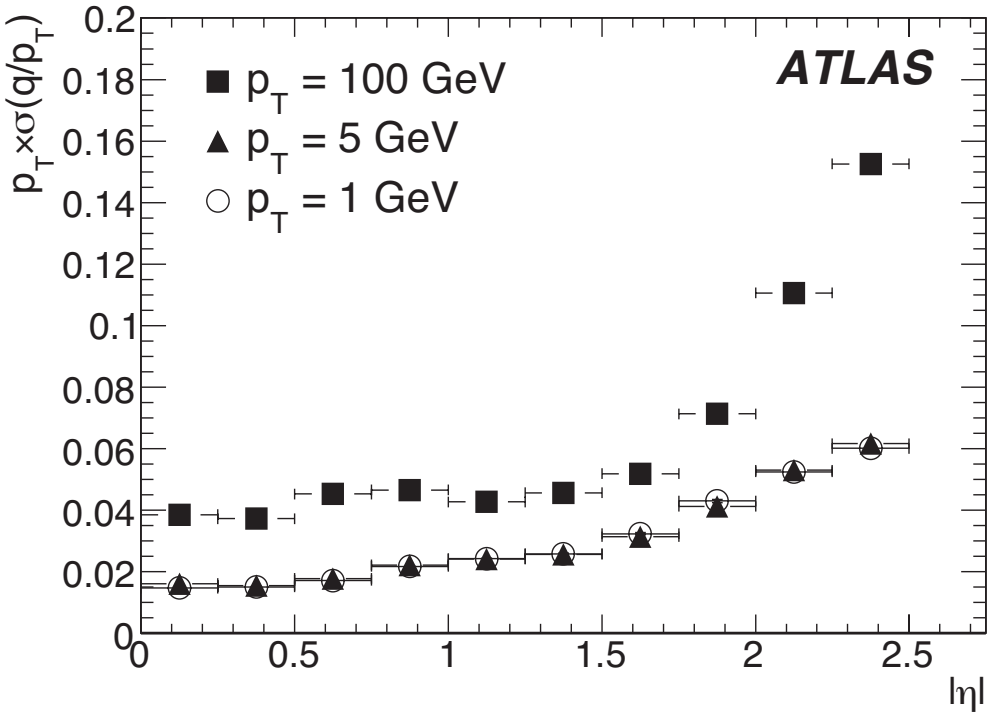
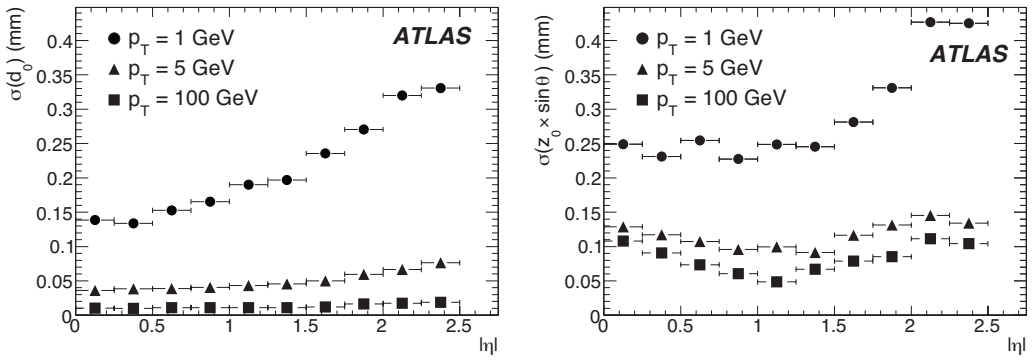



Figure 2.7: Relative $p_T$ resolution of tracks constructed using hits from the ATLAS inner detector as a function of $|\eta|$. [30]

to within $[-5, -10]$°C.

The final subsystem in the inner detector is the TRT. The TRT uses 2mm-radius drift-tubes rather than the semiconductor sensors in the SCT and pixel detector. These tubes are substantially larger than the silicon strips, 0.5-1.1m long, and a (comparatively) small number are needed to provide full coverage in the inner detector, just 350,000. The drift-tubes are bundled and arranged parallel to the beam-line in the barrel and transverse in the end-cap such that an average charged particle would leave ionisation energy in 36 tubes. The resolution of hits measured in the TRT is around 130 $\mu m$. The TRT is also capable of identifying electrons and pions based on the number of high-energy deposits left by the charged particles.

Particle hits from all parts of the inner detector are used when reconstructing tracks for later analysis. The high granularity across all detector subsytems gives accurate measurement of charged particle momentum, within the design target of $\sigma_{p_T}/p_T = 0.05 p_T \oplus 1\%$. Measurements of the $p_T$ resolution are shown in figure 2.6 and measurements of the track impact parameter resolution in figure 2.8.




(a) $d_0$ (b) $z_0$

Figure 2.8: Impact parameter resolution for tracks constructed using hits from the ATLAS inner detector as a function of $|\eta|$. [30]

### 2.2.4 Calorimetry

The ATLAS detector contains 6 separate calorimetry systems to provide coverage for both electromagnetic and hadronic particles for $|\eta| < 4.9$. Lead/liquid-Argon calorimeters are used for both the barrel and end-cap EM calorimeters, but only the end-cap hadronic calorimeters (HEC). Hadronic calorimetry in the barrel is provided by the Tile calorimeter, made from steel absorbers and plastic scintillator tiles. A dedicated copper-tungsten/liquid-Argon calorimeter called the Forward Calorimeter (FCal) provides coverage in the forward region with $|\eta| > 3.0$. Finally, the EM calorimeters are supported by a presampler in the region $|\eta| < 1.8$. A cross-section of the calorimeter system is shown in figure 2.9 and the depth and pseudorapidity coverage of each calorimeter is shown in figure 2.10.

The EM calorimeters in both the end-cap and barrel share the accordion geometry shown in figure 2.11, which gives complete coverage in $\phi$. The barrel geometry is split into two 3.2m long half-barrels covering $0 < \eta < 1.475$ and $-1.475 < \eta < 0$ and these are further split into three layers. At $\eta = 0$, layer 1 is $2.6X_0$ (electromagnetic radiation lengths) thick, layer 2 is $16X_0$ thick and the final layer is $2X_0$ thick. Depending on the pseudorapidity, the total thickness of the module is between $22X_0$ and $33X_0$. The presampler is simply a liquid-Argon layer (with no corresponding absorber) placed in front of the EM calorimeter. The two end-cap calorimeters are wheels covering the region $1.375 < |\eta| < 3.2$ and are constructed from eight wedge-shape modules with an accordion geometry. The precision region $1.5 < |\eta| < 2.5$ is instrumented with three layers.

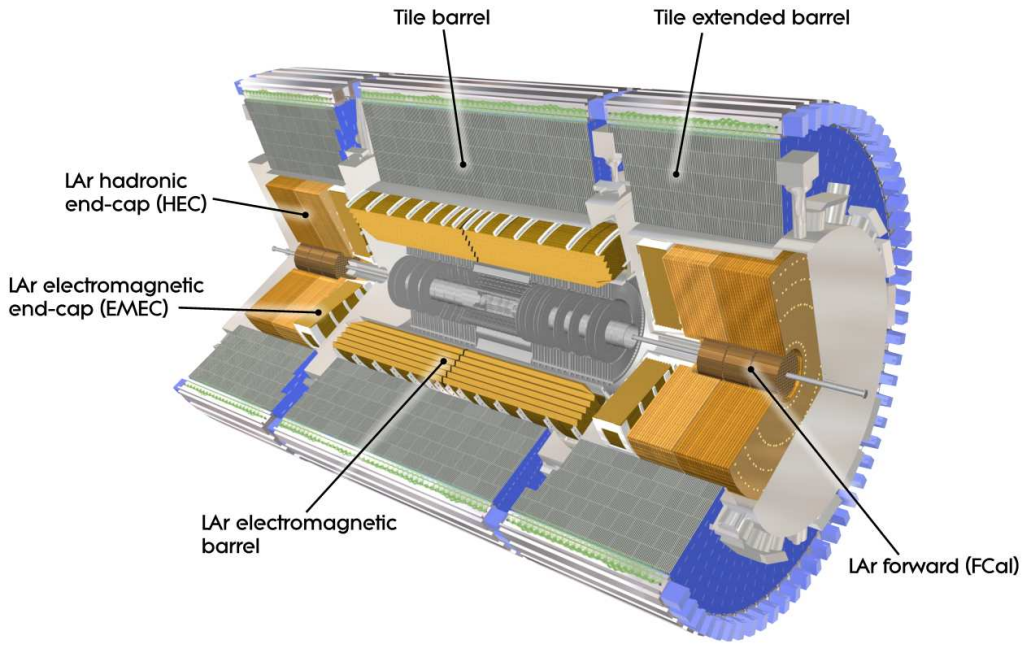



Figure 2.9: The ATLAS calorimeter system [27]. Note that beyond the components labelled in the figure, the HEC1 and HEC2 separation is visible in each end-cap.

The hadronic end-cap calorimeter (HEC) also makes use of liquid-Argon as a scintillator, but it switches to copper for the absorber; it also has a significantly different mechanical structure. Each HEC consists of an inner and outer wheel (HEC1 and HEC2 respectively) constructed of alternating copper plates and instrumented scintillator. HEC1 uses 25mm thick copper plates whereas HEC2 uses 50mm plates. The liquid Argon gap is 8.5mm for each.

Finally, there is the Tile calorimeter which surrounds both barrel and end-caps, covering the region $|\eta| < 1.7$. Unlike the liquid Argon calorimeters, this does not require active cooling, which reduced the cost and complexity of the overall calorimeter system compared to (for example) an all-LAr calorimeter. It uses 3mm thick scintillator tiles separated by 14mm steel plates to provide a high sampling frequency within a (relatively) compact space. Both the barrel and extended Tile calorimeters are segmented into three longitudinal layers with a total depth of $7.4\lambda$ (hadronic interaction lengths).

The energy resolution of a calorimeter is described by the equation:

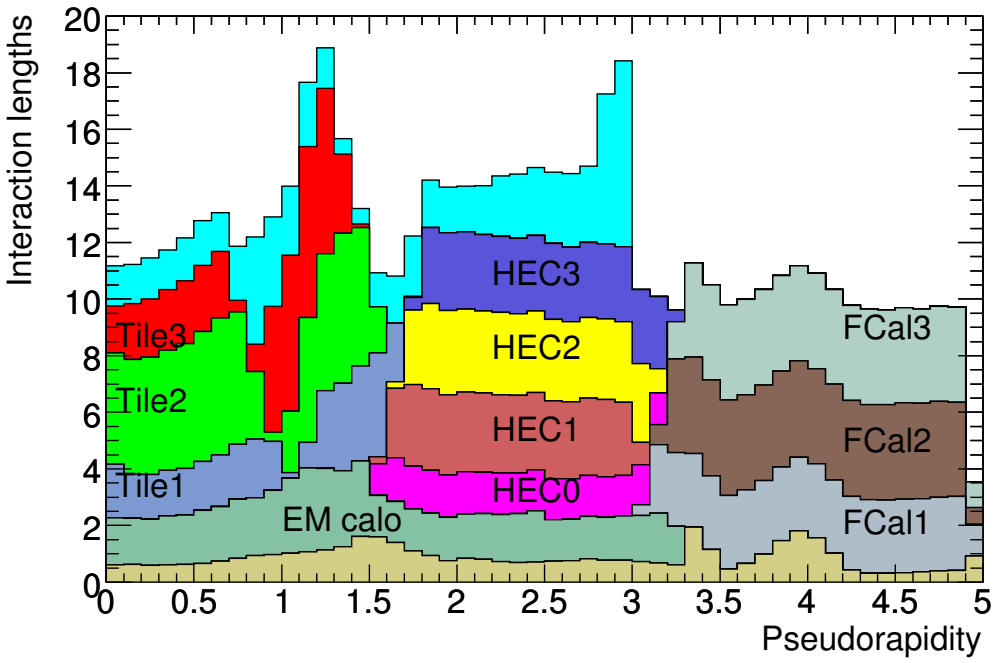



Figure 2.10: The total amount of material, in hadronic interaction lengths, in front of the calorimeters (pale brown), in the full calorimetry system, and in the muon spectrometer (pale blue) as a function of $|\eta|$. Figure from [27].

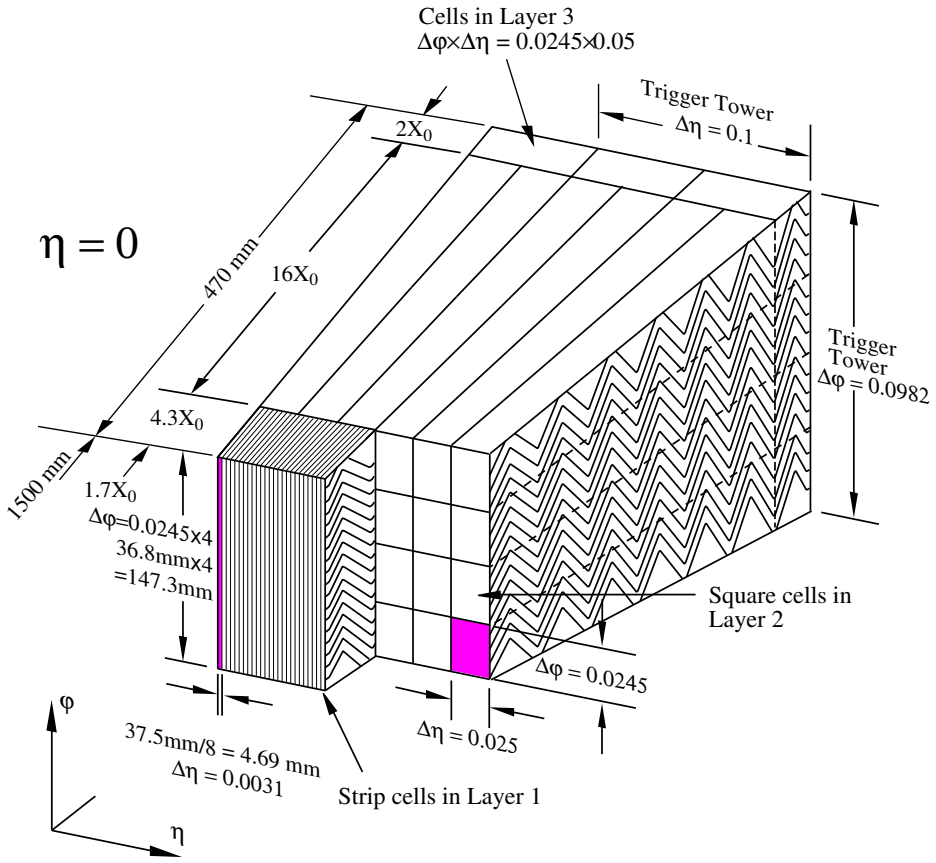



Figure 2.11: A cross-section of the EM barrel calorimeter with the accordian geometry clearly visible [27].

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.4}$$

where $a$ is the ''stochastic'' or ''sampling'' term describing fluctuations in the development of showers within the calorimeter, $b$ is the contribution from electronic noise in the readout chain, and $c$ is the constant term containing any additional behaviours which cause smearing of the measurement but do not depend on particle energy, e.g. detector non-uniformities or radiation damage. For ATLAS, the sampling and constant terms dominate but the design target for each term differs between detectors. The required resolution is $\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$ for the EM calorimeter, $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ for the hadronic barrel and end-cap calorimeters, and $\sigma_E/E = 100\%/\sqrt{E} \oplus 7\%$ for the forward calorimeters.

### 2.2.5 Muon Spectrometer

The muons created in LHC collisions will typically pass straight through the calorimeter systems and are instead measured by a dedicated muon spectrometer built around the air-core toroids that has a $p_T$ resolution of $\sigma_{p_T}/p_T = 10\%$ for muons with $p_T = 1$ TeV. The muon spectrometer (MS), shown in figure 2.12, provides coverage up to $|\eta| < 2.7$ and uses four different types of detector to deliver both precision measurements and online trigger capabilities (up to $|\eta| < 2.4$): thin gap chambers (TGC), cathode strip chambers (CSC), resistive-plate chambers (RPC), and monitored drift tubes (MDT). The MDT provide precision measurements of the muon tracks in both the end-caps and barrel and the CSC are used in the central region of the MS for accurate coordinate measurement at high rates.

High-speed tracking is also required to trigger on muons; this is provided by the RPC and TGC. The RPC are arranged in three concentric cylinders within the barrel, and the TGC are arranged as two large wheels in the end-caps. The RPC and TGC also provide additional hits for offline track reconstruction.

### 2.2.6 The ATLAS Trigger System

Given the LHC's typical collision rate of 40MHz and an average event data size of 1.5Mb, it would be challenging to even read-out the detector systems for each collision, let alone store them all. The ATLAS trigger system analyses the collisions which take place in close-to-real-time and decides which events should be saved.

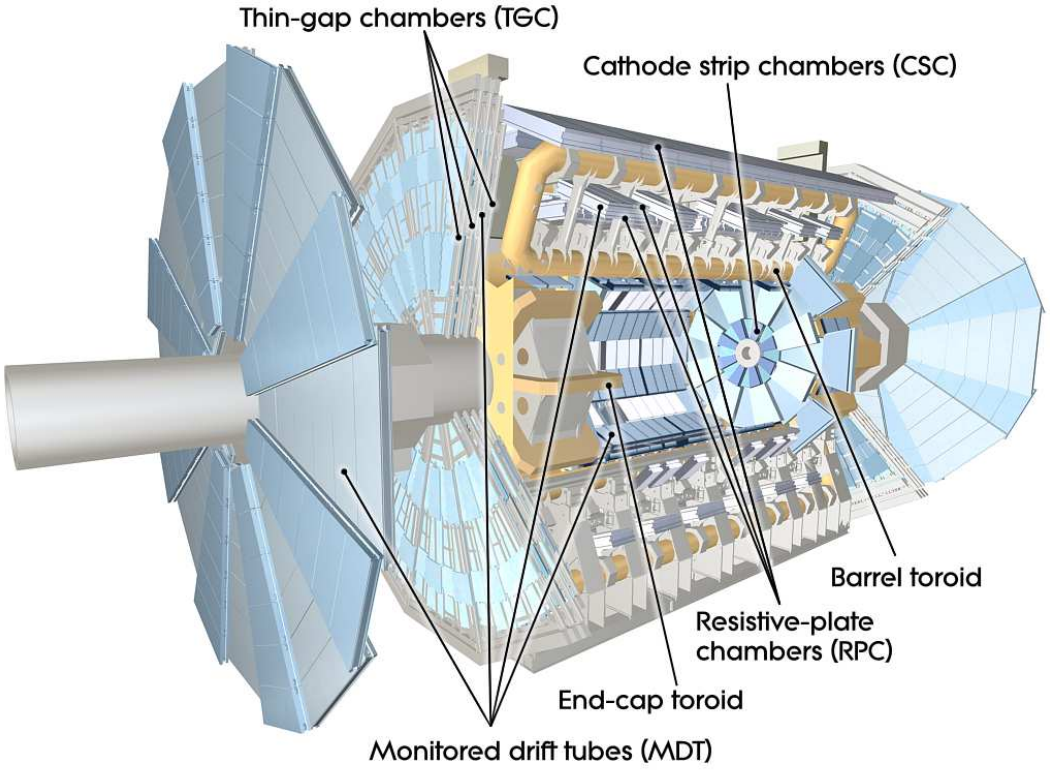



Figure 2.12: The ATLAS muon spectrometer layout. Figure from [27].

To operate within the experiment's current technical and logistical constraints, the total rate must be reduced to around 1kHz. The trigger has a two-stage architecture. The L1 (Level 1) trigger uses dedicated hardware (primarily FPGAs and ASICs) to execute simple filtering algorithms and reduce the rate from 40MHz to 100kHz. Events which are accepted by the L1 trigger are then analysed by the software-based High-Level Trigger (HLT) which further reduces the rate to 1kHz. Events which pass the HLT are then stored for later reconstruction. A simplified schematic of the trigger architecture is shown in figure 2.13. The original ATLAS trigger design specified a three level trigger, where an additional software stage ran simplified algorithms on restricted regions-of-interest before a final software stage performed whole-event reconstruction. This was replaced with a single, more flexible, software stage in which individual chains can decide at when (if ever) to access the full detector readout or use more expensive reconstruction algorithms.

The L1 trigger decision is made by the central trigger processor (CTP), which takes three inputs, the L1-Calo, L1-Muon, and L1Topo. The L1-Calo receives low-granularity information from the EM and hadronic calorimeters and uses it to identify high-pT electrons, photons, taus, jets and $E_T^{\text{miss}}$ candidates. The L1-Muon uses the RPC and TGC to determine muon candidates. In addition to the CTP, these candidates are also forwarded to the L1Topo processor. The L1Topo was introduced for run-2 and performs simple topological selections on the L1 objects, e.g. finding two electron candidates with a minimum separation in $\Delta R$. All of this information is passed to the CTP, and it makes a decision on each event based on a menu of L1 triggers. If an event is selected by any trigger from this menu then an L1-accept bit is set, and the detector data is accessed, alongside identified Regions-of-Interest (RoIs), for use in the HLT and (potentially) storage. The whole L1 decision takes place within $2.5\mu s$, of which $1.7\mu s$ is dedicated to reading out the data.

The HLT runs on a dedicated computer farm close to the ATLAS cavern. It receives the RoIs and L1-accept bits from the L1 trigger and has access to the full detector information, though for efficiency information is not read-out unless specifically requested. The HLT is built around the concept of chains. Each chain is a sequence of feature extraction (FEX) and hypothesis algorithms, where the FEX algorithms retrieve/calculate additional information about the event and hypothesis algorithms use that information to make a pass/fail decision. In the case of a failed hypothesis algorithm, the chain will cease processing immediately. Otherwise, it will continue until the final hypothesis algorithm is reached. The event is considered to have passed the chain if this last hypothesis algorithm is passed. Chains are seeded by

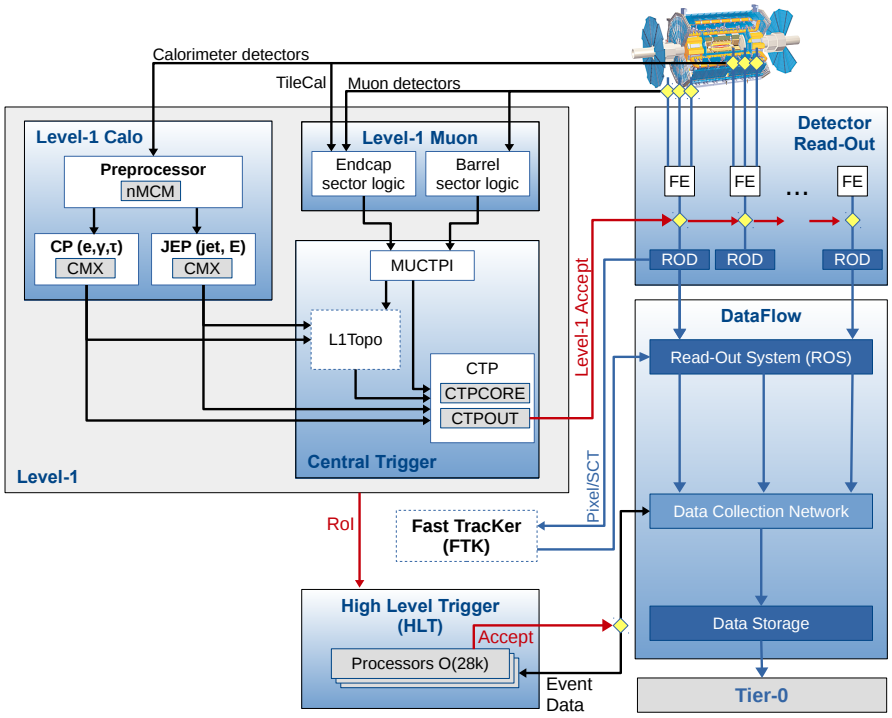



Figure 2.13: Architecture of the ATLAS trigger system during run-2. [19]

an L1 trigger and will only run when that trigger fired. The trigger menu is a set of chains which are executed in parallel (with some caching of FEXs). The processing continues until either all chains have failed or one passes. If a chain passes then the event is accepted by the trigger, and the data is read-out to storage at tier-0. The average time taken to process an event is 0.3s.

The trigger menu is chosen to maximise the possible physics studies on the data while remaining within the 1kHz output limit and contains several hundred chains. Chains can be designed to find specific physics objects, such as isolated electrons with $p_T > 100$ GeV, which may be useful for many physics analyses or a more complex combination of objects targeting a particular final state. Triggers whose rate would ordinarily be too high to store, such as low $p_T$ jet or lepton triggers, may be prescaled. An otherwise passing chain with a prescale of 100 would be accepted with a chance of $\frac{1}{100}$. Since the luminosity delivered by the LHC can vary (both during and between runs), several menus are defined, containing different chains and prescales, which can be switched between during data-taking.

# Chapter 3

# The Standard Model and Beyond

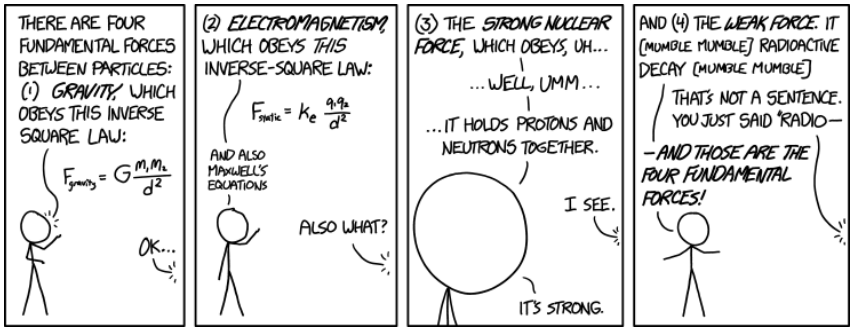



Figure 3.1: ''Of these four forces, there's one we don't really understand.'' ''Is it the weak force or the strong–'', ''It's gravity.'' *xkcd: Fundamental Forces* [31]

Even the most advanced physics experiment would be useless without some method of interpreting its results. Mathematical models provide the lens through which we can understand data, and conversely, data is the tool by which theories can be validated or refuted. In this regard, the Standard Model of particle physics is surely amongst the most successful theories in modern physics, as it provides significant explanatory power for a (relatively) small number of parameters. As shown in figures 3.2 and 3.3, the Standard Model's predictions remain accurate across many processes and orders of magnitude, and, with the exception of the discovery of neutrino masses, searches for additional particles or interactions since its completion in the 1980s have so far proven fruitless (despite significant effort).

This, however, is a thesis submitted on the topic of *experimental* particle physics

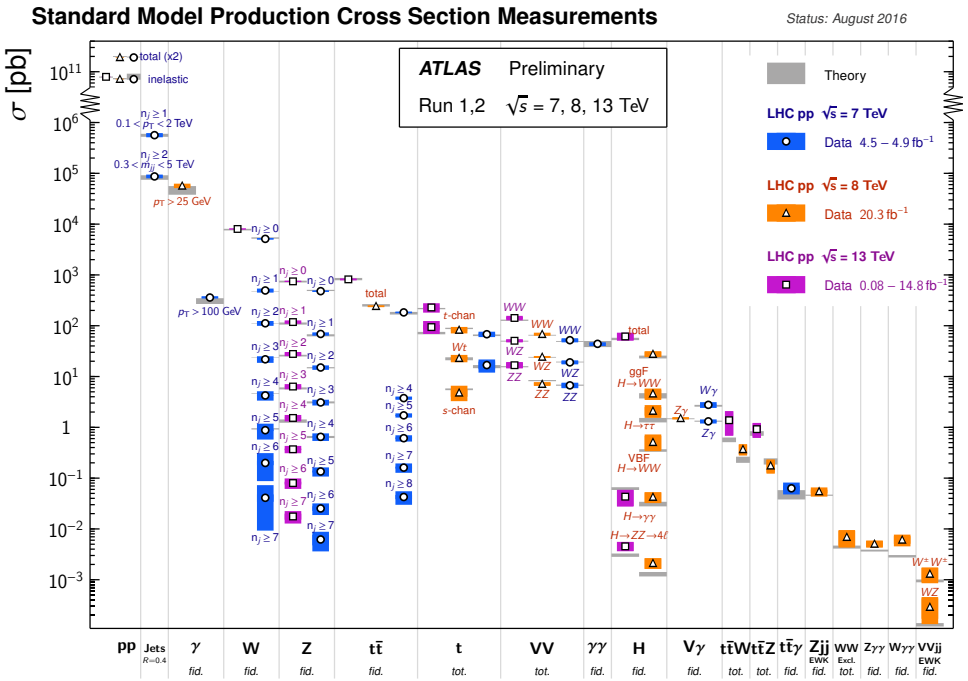



Figure 3.2: Measurements of Standard Model cross sections by the ATLAS collaboration at 7,8 and 13 TeV compared to theoretical predictions [37].

and as such the theoretical discussion will be brief. This chapter will summarise the Standard Model: its particle content and interactions, electroweak symmetry breaking, and its flaws, and present a small selection of proposed theories that extend it. Unless otherwise referenced, the content in this chapter is derived from [32–36].

## 3.1 Symmetries, Fields and Forces

The Standard Model is a relativistic quantum field theory that describes all currently known particles, their properties and their interactions.[1] These properties are entirely encoded by the Standard Model Lagrangian, $\mathcal{L}_{SM}$, that contains kinematic and interaction terms for each particle.

The construction of this Lagrangian has been guided by symmetry. Firstly as QFT is

[1]Except neutrinos, which are included but incorrectly taken to be massless. The Standard Model can be trivially modified to include massive neutrinos, but as yet there is not enough experimental evidence to decide which of several consistent modifications can be made.

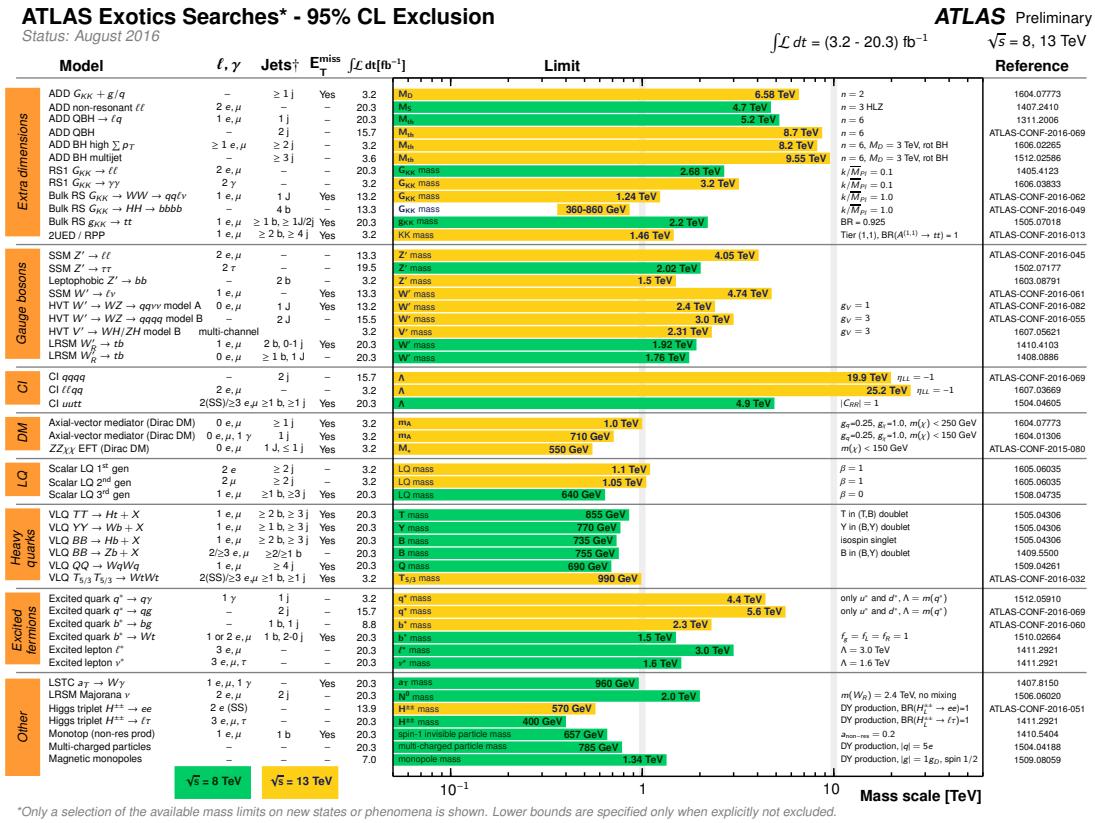



Figure 3.3: A summary of the published ATLAS limits on a wide variety of BSM (Beyond the Standard Model) processes as of August 2016 (excluding supersymmetry) [38].

relativistic, the Lagrangian must be invariant under Lorentz transformations. This requirement forces fields in QFT to exist in representations of the Lorentz group $SO^+(3,1)$. The Lorentz representation of each field fixes the spin properties of its associated particles, as shown in table 3.1. All particles in the Standard Model have spin equal to 0, 1 or $\frac{1}{2}$ but some extensions to the Standard Model also include spin-2 particles such as the graviton.

| Representation | Transformation | Spin | Particles |
|---|---|---|---|
| Scalar | $\Phi \to \Phi' = \Phi$ | 0 | $H$ |
| Vector | $V^\mu \to V'^\mu = \Lambda^\mu_\nu V^\nu$ | 1 | $W^\pm, Z^0, \gamma, G$ |
| Tensor | $T^{\mu\nu} \to T'^{\mu\nu} = \Lambda^\mu_\rho \Lambda^\nu_\sigma T^{\rho\sigma}$ | 2 | Graviton |
| Dirac | $\Psi \to \Psi' = \exp\left(\frac{1}{8}\theta_{\alpha\beta}\left[\gamma^\alpha, \gamma^\beta\right]\right)\Psi$ | $\frac{1}{2}$ | $e, \nu_e, u, d, \ldots$ |

Table 3.1: Commonly used representations of the Lorentz group and their corresponding field transformations and spin. The $\theta_{\alpha\beta}$ parameter in the Dirac representation is simply a continuous parameter of the transformation such that the field is left unchanged when all elements of $\theta$ are zero.

Particles with a spin of 1/2 are found in the Dirac representation and a unique property of these fields is that they are actually a product of two representations, which decouple if the field is massless. These representations correspond to the chirality of the field (and hence the helicity in the case of a massless particle).

This Lorentz symmetry is a *global* symmetry of the Standard Model; its transformations do not depend on the space-time coordinates of any particle. However, it was realised in the 1920s that requiring the Lagrangian to be invariant under a *local* $U(1)$ gauge symmetry generated a field equivalent to the photon field [39]. This later became the basis for Feynman's QED. Moving from $U(1)$ to higher-dimensional symmetries, a local $SU(N)$ symmetry gives rise to a force propagated by $N^2 - 1$ massless gauge bosons, as will be shown in §3.2.1.

The Standard Model of particle physics proposes that particle interactions are mediated by two forces, the strong force, which is obtained by imposing an $SU(3)_{\text{colour}}$ symmetry, and the electro-weak force, which comes from an $SU(2)_L \times U(1)_Y$ symmetry broken by the Higgs mechanism to $U(1)_{EM}$. The subscript $L$ indicates that that only left-handed chiral fermions transform non-trivially under the $SU(2)$ group. Hence the full local symmetry group of the Standard Model is:

$$SU(3)_C \times SU(2)_L \times U(1)_Y \underset{Higgs}{\to} SU(3)_C \times U(1)_{EM} \tag{3.1}$$

| Fermion | $SO^+(3,1)$ | $U(1)_Y$ | $SU(2)_L$ | $SU(3)_c$ |
|---|---|---|---|---|
| $L_L$ | $(\frac{1}{2},0)$ | $-1$ | $\mathbf{2}$ | $\mathbf{1}$ |
| $e_R$ | $(0,\frac{1}{2})$ | $2$ | $\mathbf{1}$ | $\mathbf{1}$ |
| $Q_L$ | $(\frac{1}{2},0)$ | $\frac{1}{3}$ | $\mathbf{2}$ | $\mathbf{3}$ |
| $u_R$ | $(0,\frac{1}{2})$ | $-\frac{4}{3}$ | $\mathbf{1}$ | $\bar{\mathbf{3}}$ |
| $d_R$ | $(0,\frac{1}{2})$ | $\frac{2}{3}$ | $\mathbf{1}$ | $\bar{\mathbf{3}}$ |
| Boson | | | | |
| $B$ | $(\frac{1}{2},\frac{1}{2})$ | $1$ | $\mathbf{1}$ | $\mathbf{1}$ |
| $W$ | $(\frac{1}{2},\frac{1}{2})$ | $0$ | $\mathbf{3}$ | $\mathbf{1}$ |
| $G$ | $(\frac{1}{2},\frac{1}{2})$ | $0$ | $\mathbf{1}$ | $\mathbf{8}$ |
| $h$ | $(0,0)$ | $0$ | $\mathbf{2}$ | $\mathbf{1}$ |

Table 3.2: The particle content of the Standard Model and their group representations before spontaneous symmetry breaking. $U(1)_Y$ only possesses singlet representations but as a particle's coupling to its B boson may vary this is given instead.

The full particle content of the Standard Model can then be filled out as the vector (gauge) bosons of the aforementioned symmetries and the scalar Higgs boson:

$$G, B, W, h \tag{3.2}$$

plus 3 generations of chiral fermions:

$$L_L = \begin{pmatrix} \nu_{l,L} \\ l_L \end{pmatrix}, l_R, \quad l = \{e, \mu, \tau\},$$

$$Q_L = \begin{pmatrix} q^d_L \\ q^u_L \end{pmatrix}, q^d_R, q^u_R, \quad q^d = \{d, s, b\}, q^u = \{u, c, t\}.$$

The left-handed fermions form a fundamental doublet representation under $SU(2)_L$, coupling the leptons and neutrinos, and the up and down quarks, whereas the right-handed fermions exist in separate $SU(2)_L$ singlet states. Their representations under each symmetry group are given in table 3.2. Note that the quarks also exist in a colour triplet representation

$$q = \begin{pmatrix} q_r \\ q_g \\ q_b \end{pmatrix} \tag{3.3}$$

## 3.2 Quantum Chromodynamics

### 3.2.1 The $SU(3)$ Lagrangian

The strong force between quarks and gluons is described by Quantum Chromodynamics (QCD), an $SU(3)$ Yang-Mills theory. That $SU(3)$ is the unique choice for describing QCD can be seen by considering several experimental results. Firstly, only colour-singlet mesons and baryons are observed in nature, so there must be some mechanism for disallowing the observation of coloured combinations of quarks. This can be explained by the confinement property of $SU(N)$ gauge theories as will be discussed in §3.2.2 (in fact, only $SU(N)$ theories possess this property[2]). Secondly, some measurements depend strongly on the number of quark colours, for example, $ee \to hadrons$ production should (at tree level) be proportional to $ee \to \mu\mu$ with additional multiplicative factors for the number of colours, $N_C$, and flavours.

$$\frac{\sigma(ee \to hadrons)}{\sigma(ee \to \mu\mu)} \approx N_C \sum_f Q_f^2 \tag{3.4}$$

where $f$ are the quark flavours and $Q_f$ is the quark's electric charge.

To match experimental data $N_C$ must be equal to 3 and the only $SU(N)$ theories with triplet representations are $SU(2)$ and $SU(3)$. The $SU(2)$ triplet representation is real so the quarks and anti-quarks would have an equal colour charge, however, the observation of colourless $q\bar{q}$ mesons but not $qq$ or $\bar{q}\bar{q}$ ones suggests that this cannot be the case. Quarks must exist in a complex representation. Hence, $SU(3)$ is the only possible Yang-Mills theory consistent with experiment.

While the mechanics of QCD calculations are beyond the scope of this thesis, it is still instructive to derive its mathematical structure and some general properties of Yang-Mills theories.

Consider the free Lagrangian of an $SU(3)$ colour triplet

$$\Psi = \begin{pmatrix} \psi_r \\ \psi_g \\ \psi_b \end{pmatrix} \tag{3.5}$$

[2]This includes the unbroken electroweak group $SU(2)_L$ but following symmetry breaking confinement no longer occurs. This can be seen intuitively by considering that QCD confinement can be attributed to the existence of *massless* and *self-interacting* gluons. While the electroweak W and Z bosons are self-interacting, their large masses prevent confining behaviour.

where $\Psi$ consists of three mass-degenerate quark fields each with a ''colour'' charge labelled $r, g, b$ (or $\bar{r}, \bar{g}, \bar{b}$). $\mathcal{L}_{\text{free}}$ would be

$$\mathcal{L}_{\text{free}} = \bar{\Psi} i\gamma^{\mu}\partial_{\mu}\Psi - m\bar{\Psi}\Psi. \tag{3.6}$$

This is invariant under the global $SU(3)$ transformation $\Psi \to \Psi' = U\Psi$.

$$U = e^{-i\frac{\lambda_a}{2}\theta_a} \tag{3.7}$$

where $\lambda_a$ are the Gell-Mann matrices, the generators of $SU(3)$, and $\theta_a$ are the transformation parameters. If $\Psi$ transforms under a local gauge transformation (ie. $\theta = \theta(\mathbf{x})$) then $\mathcal{L}_{\text{free}}$ is no longer invariant and the definition of $\partial_{\mu}\Psi$ no longer has a geometric interpretation

$$\begin{aligned} \partial_{\mu}\Psi &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\Psi(x+\epsilon n_{\mu}) - \Psi(x)\right] \\ &\to \left[U(x+\epsilon n_{\mu})\Psi(x+\epsilon n_{\mu}) - U(x)\Psi(x)\right] \neq U(x)\partial_{\mu}\Psi. \end{aligned} \tag{3.8}$$

Sensible behaviour of the derivative operator is preserved by replacing the derivative $\partial_{\mu}$ with the covariant derivative

$$D_{\mu} = \partial_{\mu} + ig_s\lambda_{\alpha}G^{\alpha}_{\mu} \tag{3.9}$$

where $G^{\alpha}_{\mu}$ are the gluon fields, and $g_s$ is the strong coupling constant. The gluon fields couple the local symmetries at each point in space-time and restore $SU(3)$ symmetry under both local and global transformations. In fact, invariance of any local $SU(N)$ symmetry *requires* the existence of such a vector field for each of its generators. The gluon fields make it possible to construct one final gauge invariant term. First, define the field strength tensor $F^{a}_{\mu\nu}$

$$F^{a}_{\mu\nu} = \partial_{\mu}G^{a}_{\nu} - \partial_{\nu}G^{a}_{\mu} - g_s f^{a}_{bc}G^{b}_{\mu}G^{c}_{\nu} \tag{3.10}$$

$$F^{a}_{\mu\nu} \to F'^{a}_{\mu\nu} = U(\mathbf{x})F_{\mu\nu}U^{-1}(\mathbf{x}) \tag{3.11}$$

and note that $F^{a\dagger}_{\mu\nu}F^{\mu\nu}_{a}$ is a singlet under $SU(3)$. This is the only additional term we can add to the Lagrangian which preserves the $SU(3)$ symmetry, CP symmetry[3] and is of $dim \leq 4$. This term is added to give the full Yang-Mills Lagrangian for an

[3]There is one additional allowed term under $SU(3)$ which is proportional to $\theta\epsilon^{\mu\nu\rho\sigma}F^{a}_{\mu\nu}F^{b}_{\rho\sigma}$ and violates CP symmetry, but $\theta$ has been found to be $\ll 1$. The ''unnaturalness'' of this is known as the strong CP problem. There have been several proposed solutions, of which the most well known is the axion.

$SU(3)$ symmetry

$$\mathcal{L}_{SU(3)} = \bar{\Psi} i\gamma^\mu D_\mu \Psi - m\bar{\Psi}\Psi - \frac{1}{4g_s^2} F^{a\dagger}_{\mu\nu} F_a^{\mu\nu} \tag{3.12}$$

Nothing (aside from the labelling) in this derivation is unique to $SU(3)$ and two general properties of this Lagrangian should be stressed. Firstly, the expanded field strength term includes 3 and 4-field self-couplings ($(\partial_\mu G^a_\nu)^\dagger G^\mu_b G^\nu_c$ and $(G^a_\mu G^b_\nu)^\dagger G^\mu_c G^\nu_d$ respectively) so the gauge fields of any non-abelian[4] symmetry are self-interacting. Secondly, there is no possible mass term for these gauge fields.

Keeping these facts in mind, it is now possible to introduce the full QCD Lagrangian, which includes all 6 flavours of quark.

$$\mathcal{L}_{QCD} = -\frac{1}{4g_s^2} F^{a\dagger}_{\mu\nu} F_a^{\mu\nu} + \sum_{f \in \text{flavours}} \bar{q}_{f,\alpha} (i\gamma_\mu D_\mu)_{\alpha\beta} q_{f,\beta} \tag{3.13}$$

Where $\alpha, \beta \in \{1, 2, 3\}$ are the colour indices. Note that the mass terms have been dropped; while these are perfectly valid under QCD, due to the chiral nature of the fields they break electroweak symmetry as will be shown in §3.3.1.

### 3.2.2 Confinement and Asymptotic Freedom

It was highlighted in the introduction to this section that QCD possesses the peculiar property of *confinement*, whereby the direct observation of colour-charged particles is not possible. This is understood as the effective strength of the strong force between two colour-charged particles growing with separation. Eventually, the potential energy between them is so great that it is energetically favourable to create a new particle pair. Unfortunately, this observation leaves further theoretical predictions in a difficult situation. Since the coupling is large ($\alpha_s \approx \mathcal{O}(1)$), using a perturbative approach is not viable. Some calculations can be made using the lattice approach, but this is computationally expensive.

However, there is a second experimental observation that seemingly contradicts this view. In the interaction of high-energy hadrons, quarks seem to behave as free particles and their interactions can be predicted extremely accurately with a perturbative parton model. In the 1970s theorists realised that not only could

[4]Defining $\mathbb{G}_\mu = \lambda_a G^a_\mu$ shines some light on this property. Under this definition $F_{\mu\nu}$ becomes $\mathbb{F}_{\mu\nu} = \partial_\mu \mathbb{G}_\nu + \partial_\nu \mathbb{G}_\mu + i g_s [\mathbb{G}_\mu, \mathbb{G}_\nu]$. Gauge field self-coupling requires the final term and this is obviously zero for any abelian symmetry group (such as the $U(1)$ of QED) and non-zero for any non-abelian symmetry.

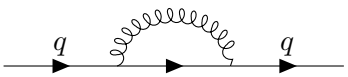



Figure 3.4: The 1-loop quark propagator.

perturbative calculations prove accurate at higher-energies, they also provided a rationale for the change in coupling strength. It is now known that as the energy of interaction increases (and the length scale becomes very short) the apparent QCD coupling decreases and approaches a free theory, a property called *Asymptotic Freedom*. The point at which the theory becomes accessible to perturbative methods, i.e. $g \sim 1$, is known as the scale of QCD, $\Lambda_{QCD}$. Its exact value depends on the renormalisation scheme in which one works, but in the modified minimal subtraction scheme $\Lambda_{QCD}^{\overline{MS}} = 217 \pm 24$ MeV.

In fact, the transition from asymptotic freedom to confinement follows directly from the QCD Feynman rules calculated in the perturbative regime. Consider the NLO contribution to the quark self-coupling, shown in figure 3.4; the momentum of the gluon propagator is unconstrained by the incoming and outgoing quark momentum leading to a logarithmic term

$$\int \frac{d^4k}{k^4} \propto \log k \tag{3.14}$$

that blows up as $k$ become large. This is a ubiquitous occurrence in quantum field theories, with a well-established method for its solution, *renormalisation*. First, some method is used to regularise the integral, either by introducing a momentum cutoff $\Lambda$ or shifting the integral from $4 \to 4 + \epsilon$ dimensions where the divergent component is controlled by $\epsilon$. These methods introduce new terms proportional to

$$\alpha_s \log\left(\frac{\Lambda}{\mu}\right) \tag{3.15}$$

(and higher powers thereof) where $\mu$ is an arbitrary mass scale. Provided there are only a finite number of such integrals (a requirement fulfilled by $SU(N)$ theories whose Lagrangian has mass-dimension equal to 4) an equal number of counter-terms can be added to the Lagrangian to remove these divergences. If this is not the case, as will be seen for field theories of gravity in §3.4.1, then an infinite number of counter-terms are required, and the theory is no longer predictive. In any physical result, the dependence on $\Lambda$ must cancel, but the dependence on $\mu$ remains, typically as the ratio $\frac{Q}{\mu}$ where $Q$ is the energy scale of the process. These terms are absorbed into the bare coupling constant that becomes a scale-dependent physical coupling. The equation describing this change is known as the $\beta$ function.

The scale-dependent behaviour of the QCD physical coupling is governed by the QCD $\beta$ function

$$\beta(\alpha_s) = \mu^2 \frac{\partial \alpha_s}{\partial \mu^2} = -b_0 \alpha_s^2 + b_1 \alpha_s^3 + O(\alpha_s^4), \tag{3.16}$$
$$b_0 = \frac{33 - 2n_f}{12\pi}, b_1 = \frac{153 - 19n_f}{24\pi^2}$$

where $n_f$ is the number of flavours of quark. The negative sign in front of the $\alpha_s^2$ term gives rise to the confining behaviour at low energy. The removal of all infinities from the theory also requires that the quark masses become similarly scale-dependent.

Evaluating (3.16) (to $O(\alpha_s^2)$) at the renormalisation scale and integrating gives

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2) b_0 \log(Q^2/\mu^2)} \tag{3.17}$$

which clearly decreases as $Q$ increases. In practice, $\mu$ is often chosen to be $\Lambda_{QCD}$. The predicted running behaviour of all three forces is shown in figure 3.5, the measured behaviour of $\alpha_s$ is shown in figure 3.6.

One note of caution must be sounded about this argument for the existence of asymptotic freedom and confinement from a perturbative approach. While perturbation theory does predict a region where $\alpha_s$ becomes strong, the accuracy of perturbative methods break down as this becomes the case, and as such, it is not necessarily accurate to extrapolate the behaviour at large $\alpha_s$/low-energy from the perturbative/high-energy region.

### 3.2.3 Deep Inelastic Scattering

Confinement is certainly an interesting phenomenological feature of the strong force, but it is quite inconvenient for collider experiments. Obtaining lone quarks to collide is somewhat challenging. In practice, the LHC accelerates and collides protons and this introduces an additional level of complexity to theoretical predictions. It is not enough to be able to calculate Feynman diagrams for quarks, the structure of the two colliding protons must also be taken into account. Unfortunately, the binding of quarks within protons is firmly within the non-perturbative realm of QCD, so a first principles approach is impractical.

For a high-energy collision, we may assume the quarks and gluons within the col-

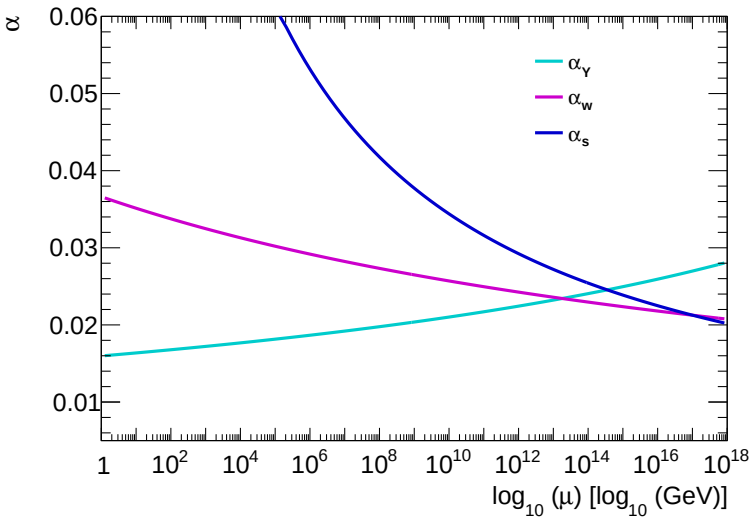
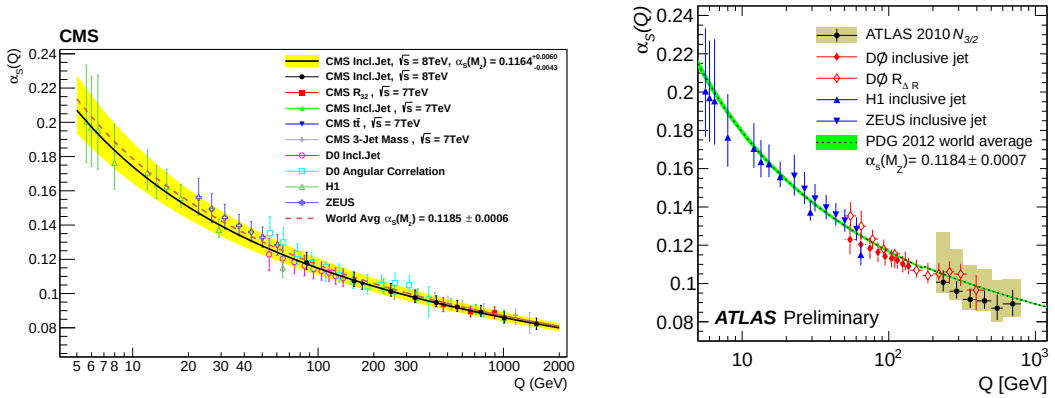



Figure 3.5: The dependence of the strong, weak and hypercharge couplings on the log of the interaction scale $\mu$ at one loop level. The values given are in terms of $\alpha = g/4\pi$ [40]. Higher-order calculations give only minor corrections to this behaviour [41].




(a) (b)

Figure 3.6: The CMS (a) [42] and ATLAS (b) [43] measurements of $\alpha_s$ with respect to the interaction scale $Q$. The CMS measurements includes a variety of processes measured at both $\sqrt{s} = 7, 8$ TeV, the ATLAS measurements are at 7 TeV only. On the ATLAS plot, the black bars indicate experimental error and the brown bars the total error. Good agreement is found between experiment and the two-loop theoretical calculation in both cases.

liding protons would be asymptotically free. Therefore we can split the theoretical predictions into two pieces, a universal proton structure function, describing the distribution of quarks and gluons within the proton, and the perturbative interaction process.[5]The first can be measured independently and then assumed to be valid for all protons. Deep inelastic scattering makes these measurements through high-energy leptonic collisions with either moving or static hadronic targets.

In the language of deep inelastic scattering, quarks and gluons are referred to as partons and each parton carries a fraction $x$ of the proton's total momentum. The probability of interaction with a parton of momentum fraction $x$ is specified for each type of parton ($g, u, d, s, c$+antiparticles) by a *parton distribution function* (PDF) $f(x, Q^2)$, where $Q^2$ is the momentum scale of the interaction. The cross-section of an interaction between two protons is then given in terms of the hard process $\hat{\sigma}(x_1, x_2, Q^2)$ and these PDFs.

$$\sigma = \int_{x_1} dx_1 f(x_1, Q^2) \int_{x_2} dx_2 f(x_2, Q^2) \hat{\sigma}(x_1, x_2, Q^2) \tag{3.18}$$

Note that this equation is not necessarily IR-safe. The IR-complete equation factorises out the low $x$ contribution and introduces a dependence on the factorization scale in a fashion similar to renormalisation but this is beyond the scope of this discussion.

Within the proton (or similar bound state) the PDFs are bound by a set of relations such that overall the bound state acts as having the expected flavour numbers, e.g. for the proton:

$$\begin{aligned}
&\int_0^1 u(x, Q^2) - \bar{u}(x, Q_2) dx = 2, && \int_0^1 d(x, Q^2) - \bar{d}(x, Q_2) dx = 1 \\
&\int_0^1 s(x, Q^2) - \bar{s}(x, Q_2) dx = 0, && \int_0^1 c(x, Q^2) - \bar{c}(x, Q_2) dx = 0 \\
&\int_0^1 g(x, Q^2) - \bar{g}(x, Q_2) dx = 0.
\end{aligned} \tag{3.19}$$

Examples of the shape of these PDFs is given for the CTEQ14 PDF set in figure 3.7.

[5]A similar argument can be made to factorise the hard interaction producing a quark or gluon and its subsequent fragmentation and hadronisation. This will be discussed in more detail in §4

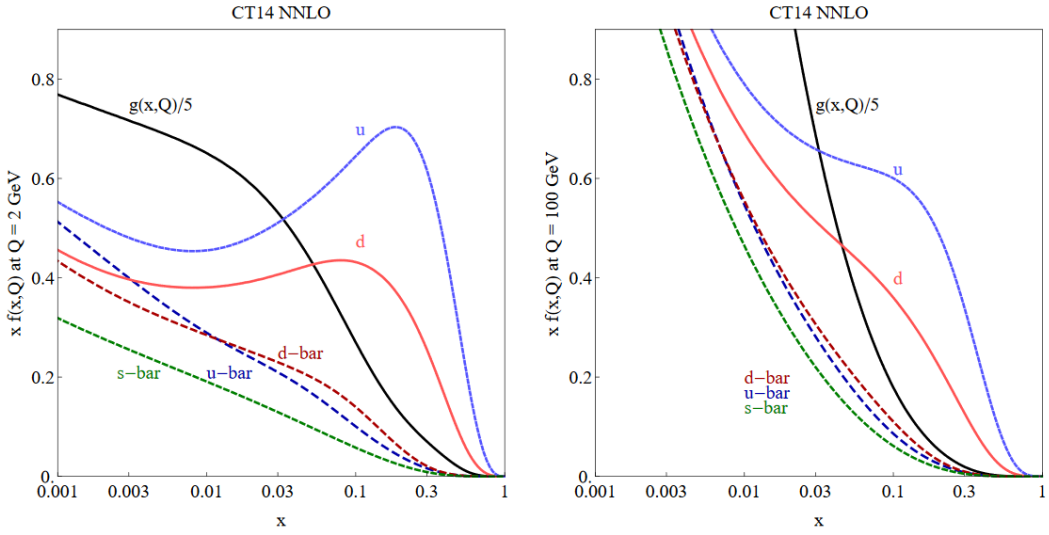



Figure 3.7: PDFs for the $u, \bar{u}, d, \bar{d}, s, \bar{s}$ quarks and gluon taken from the CT14 PDF set. Left is at $Q^2 = 2$ GeV and right at $Q^2 = 100$ GeV [44].

## 3.3 Electroweak Theory

The second piece of the Standard Model is the electroweak theory, which brings together the weak and electromagnetic forces. To motivate the following section, it is worth (briefly) examining the history of its creation. The weak force was first posited in some form in 1933 as the Fermi theory of beta decay, a simple 4-fermion contact interaction involving the previously unobserved neutrino, as an explanation for the continuous energy spectrum observed in beta decay (see figure 3.10), though the neutrino itself was not observed until 1956 [45]. At a similar time, QED was being developed and tested extensively, and it was found to give accurate predictions for previously unexplained phenomena such as the correction to the electron's magnetic moment. This lead to a great effort to explain the weak force as a similar quantum field theory and later to unify the two forces; a task made more complex by the evidence emerging in the 1950s that indicated the weak interaction had the strange property of violating parity symmetry.

There was an additional stumbling block in the development of a full quantum field theory of the weak interaction. From the lack of observation of its gauge boson(s) the interaction must be short-ranged. This required the gauge boson(s) to be massive, but any suitable mass term would break the gauge symmetry. This impasse was eventually resolved by the discovery that spontaneous symmetry breaking provided a mechanism to give mass to the gauge bosons and led to the realisation of electroweak

theory in 1969, swiftly followed by proof of its renormalisability in 1971.

### 3.3.1 Weak Interactions and Chirality

The electroweak Lagrangian is formulated as the product of two symmetries, $SU(2)_L$ which acts solely on the left-handed fermions (and right-handed antifermions), and the $U(1)_Y$ symmetry. Their charges are labelled *weak isospin* ($T$) and hypercharge ($Y$) respectively. As the weak interaction is maximally parity violating, only left-handed fermions are charged under it. However, parity is preserved under $U(1)$ interactions which affect both left and right-handed fermions equally.

With these two symmetries we can write down the $SU(2)_L \times U(1)_Y$ Lagrangian by a similar method to §3.2.1.

$$\begin{aligned}\mathcal{L}_{SU(2)\times U(1)} = &\sum_{\psi_f \in fermions} \bar{\psi}_f(\partial_\mu - \frac{ig}{2}\tau^i W^i_\mu \mathbf{P}_L - \frac{ig'}{2} y_f B_\mu)\psi_f \\ &- \frac{1}{4}B^{\mu\nu}B_{\nu\mu} - \frac{1}{4}W_i^{\mu\nu}W^i_{\nu\mu};\end{aligned} \tag{3.20}$$

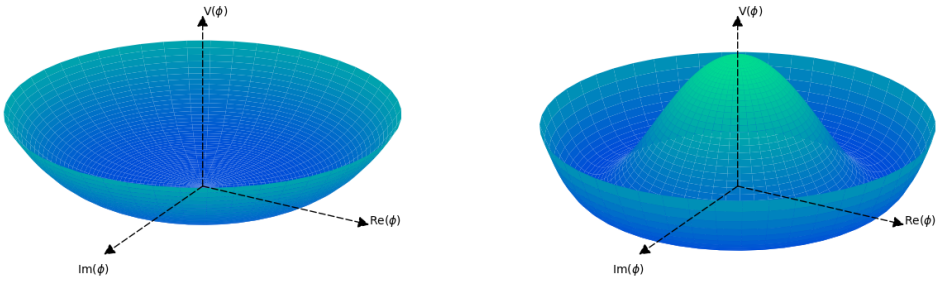
$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \qquad W^i_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - ig\epsilon^{ijk}W^j_\mu W^k_\nu$$

Here $B_\mu$ and $W^i_\mu, i \in 1,2,3$ are the vector fields associated with the $U(1)_Y$ and $SU(2)_L$ symmetries respectively and $\tau^i$ are the generators of the $SU(2)$ group. There are two other new terms in this equation, the $\mathbf{P}_L = \frac{1}{2}(1-\gamma^5)$ operator projects the left-chiral field from the full Dirac spinor. The $y_f$ term is the hypercharge for each fermion. In non-abelian $SU(N)$ theories, the coupling between the gauge and matter fields is fixed by the chosen representation of the fermions under that symmetry. For abelian symmetries this is not so; the coupling can be chosen freely for each field.

Aside from the gauge boson masses, the other obvious missing piece of this Lagrangian is a mass term for the fermions. While we were free to add this under $SU(3)_c$ the chiral nature of $SU(2)_L$ symmetry prohibits such terms.

$$m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \xrightarrow{SU(2)_L} m(U^{-1}\bar{\psi}_L\psi_R + U\bar{\psi}_R\psi_L). \tag{3.21}$$

Thankfully, the spontaneous symmetry breaking procedure will provide a solution for this as well.




(a) $\mu^2 > 0$ (b) $\mu^2 < 0$

Figure 3.8: The parabolic and ‘‘Mexican hat’’ potentials corresponding to $V(\phi) = \mu^2\phi^\dagger\phi + \frac{\lambda}{4}(\phi^\dagger\phi)^2$. $\mu^2 > 0$ (a) has a unique minimum at $v = 0$ but for $\mu^2 < 0$ (b) a continuum of degenerate minima lie along a ring at $v = \frac{2\mu^2}{\lambda}$.

### 3.3.2 Spontaneous Symmetry Breaking

So far only explicit symmetries, where the ground state preserves the symmetry of the Lagrangian, have been examined. However, it is possible to introduce a new scalar field with a potential chosen such that the ground state does not preserve this symmetry. For example, if a new complex scalar field $\phi$ has the potential $V = -\mu^2\phi^\dagger\phi + \frac{\lambda}{4}(\phi^\dagger\phi)^2$ then there is a degenerate spectrum of possible ground states, as shown in figure 3.8.

By introducing symmetry breaking in this way, $\phi$ can be expanded around a chosen minimum $v$ as $\phi = (v + h(x)) \exp[i\theta(x)/2]$ where $h(x)$ is the Higgs field associated with radial perturbations away from the vacuum minima and $\theta(x)$ are the Goldstone bosons associated with angular perturbations between different minimas. By choosing an appropriate gauge, the Goldstone bosons can be ‘‘eaten’’ by the existing gauge bosons. This provides the gauge bosons with an additional degree of freedom and allows them to acquire a mass without breaking the gauge symmetry of the Lagrangian.

In full electroweak theory, we wish to break the $SU(2)_L \times U(1)_Y$ symmetry to $U(1)_{EM}$ so we require a complex $SU(2)_L$ doublet of scalar fields. By choosing

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{3.22}$$

(with v and H real) all 4 generators are broken but the linear combination $\frac{Y}{2} + \tau_3$ is preserved. The labelling of the doublet fields as $(0, +)$ is based on their future charge under this combination. By setting $\phi_0 = v + H(x)$ , $H(x)$ is uncharged. The ungauged Higgs Lagrangian is given by

$$\mathcal{L}_{Higgs} = (D_\mu \phi)^\dagger (D^\mu \phi) + \mu^2 \phi^\dagger \phi - \frac{\lambda}{4}(\phi^\dagger \phi)^2 \tag{3.23}$$

$$D^\mu = \left( \partial^\mu - \frac{ig\tau^i}{2} W_i^\mu - \frac{ig'}{2} B^\mu \right).$$

Inserting (3.22) gives

$$D_\mu \phi = \frac{1}{\sqrt{2}} \left[ \partial_\mu \mathbb{I} - \frac{ig}{2} \begin{pmatrix} W_\mu^3 + \frac{g'}{g} B_\mu & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & W_\mu^3 - \frac{g'}{g} B_\mu \end{pmatrix} \right] \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}$$

which makes explicit the observed mixing of the $W^i$ and $B$ fields. Expanding (3.23) and only keeping the kinetic and mass terms we find

$$\mathcal{L}_{Higgs} = \frac{1}{2} \partial_\mu H \partial^\mu H - \frac{\mu^2}{2} H^2 \tag{3.24}$$

$$+ \frac{g^2 v^2}{8} (W_\mu^1 W_1^\mu + W_\mu^2 W_2^\mu) \tag{3.25}$$

$$+ \frac{g^2 v^2}{8} (W_\mu^3 - \frac{g'}{g} B_\mu)(W_3^\mu - \frac{g'}{g} B_\mu) \tag{3.26}$$

ie. we have a scalar boson with mass $\mu/\sqrt{2}$ and the $W^1, W^2$ gauge bosons have acquired a mass of $gv/2\sqrt{2}$. The $W^3$ and $B$ fields have mixed, leaving what appears to be a mass term for the combined field $(W_\mu^3 - \frac{g'}{g} B_\mu)$ but the combination $(W_\mu^3 + \frac{g'}{g} B_\mu)$ remains massless. If we introduce the *electroweak mixing angle*, $\theta_W$, then we can define two new fields:

$$Z^\mu = \cos\theta_W W_3^\mu - \sin\theta_W B^\mu \tag{3.27}$$

$$A^\mu = \sin\theta_W W_3^\mu + \cos\theta_W B^\mu \tag{3.28}$$

Here, the trigonometric relations can also be expressed in terms of the relative strength of the electromagnetic and weak coupling constants:

$$\cos\theta_W = \frac{g}{(g^2 + g'^2)^{1/2}}, \qquad \sin\theta_W = \frac{g'}{(g^2 + g'^2)^{1/2}} \tag{3.29}$$

With this redefinition, the familiar electromagnetic force emerges, propagated by the massless photon $A$ alongside the now massive $W^+, W^-$, and $Z^0$ bosons. A key experimental test of spontaneous symmetry breaking is that it predicts a precise relationship between the masses of the $W$ and $Z$ bosons

$$M_Z = \frac{M_W}{\cos\theta_W}, \tag{3.30}$$

which can be verified against independent measurements of $\theta_W$ such as from neutrino experiments [46].

### 3.3.3 Fermion Masses

Since (3.21) forbids traditional mass terms for chiral fermions under $SU(2)_L$, it falls to the Higgs mechanism to provide their masses. Consider replacing the mass parameter in (3.21) by a Higgs doublet. The transformation of the doublet should cancel that of the left-handed fermion and preserve $SU(2)_L$ symmetry. Assuming a lepton $l$ with (yukawa) coupling $y_l$, then after we gauge the doublet as in (3.22) we are left with a Higgs interaction term and a mass term for the bottom component of the $\psi_L$ doublet. For $\psi_L^T = (\,\nu_L \; l_L\,)$

$$y_f(\bar{\psi}_L\phi l_R + \bar{l}_R\phi^\dagger\psi_L) \to \frac{y_l v}{\sqrt{2}}(\bar{l}_L l_R + \bar{l}_R l_L) + \frac{y_l H}{\sqrt{2}}(\bar{l}_L l_R + \bar{l}_R l_L) \tag{3.31}$$

By rewriting $m_l = y_l v/\sqrt{2}$ this becomes

$$m_l(\bar{l}_L l_R + \bar{l}_R l_L) + \frac{m_l h}{v}(\bar{l}_L l_R + \bar{l}_R l_L). \tag{3.32}$$

The top component of the $\psi_L$ doublet can also become massive by coupling with the conjugate Higgs field

$$\phi_C = \begin{pmatrix} \phi_0* \\ -\phi_- \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} v+H \\ 0 \end{pmatrix}. \tag{3.33}$$

Thus, the Higgs field provides a mechanism for fermions to gain mass and couples to them with a strength proportional to their mass. Unfortunately, when more than one generation of fermion exists a further complication emerges. We are generally required to allow any terms compatible with the symmetry requirements of our Lagrangian, and it is possible to generalise these mass terms further. For 3 generations of chiral fermions (for now labelled $u^i, d^i, i \in \{1,2,3\}$) the *Yukawa* sector of

the Lagrangian is given by

$$\mathcal{L}_{Yukawa} = \sum_{i=1}^{3}\sum_{j=1}^{3}\left[y^d_{ij}\bar{\psi}^i_L\phi d^j_R + y^u_{ij}\bar{u}^i_R\phi^\dagger\psi^j_L\right] + h.c. \tag{3.34}$$

Collecting the $y^d_{ij}, y^u_{ij}$ parameters into two $3\times 3$ matrices $\mathcal{M}^u, \mathcal{M}^d$ the mass terms become

$$\begin{pmatrix}\bar{d}_1 & \bar{d}_2 & \bar{d}_3\end{pmatrix}_R \mathcal{M}^d \begin{pmatrix}d_1\\ d_2\\ d_3\end{pmatrix}_L + \begin{pmatrix}\bar{u}_1 & \bar{u}_2 & \bar{u}_3\end{pmatrix}_R \mathcal{M}^u \begin{pmatrix}u_1\\ u_2\\ u_3\end{pmatrix}_L \tag{3.35}$$

In the Standard Model, the quark (which this notation is obviously alluding to) and lepton mass matrices can be diagonalised, rotating the $\psi$ vectors into consistent mass eigenstates

$$\begin{pmatrix}u_1\\ u_2\\ u_3\end{pmatrix}_{L,R} = U_{L,R}\begin{pmatrix}u\\ c\\ t\end{pmatrix}_{L,R}, \quad \begin{pmatrix}d_1\\ d_2\\ d_3\end{pmatrix}_{L,R} = D_{L,R}\begin{pmatrix}d\\ s\\ b\end{pmatrix}_{L,R} \tag{3.36}$$

$$U_R^{-1}\mathcal{M}^u U_L = \begin{pmatrix}m_u & 0 & 0\\ 0 & m_c & 0\\ 0 & 0 & m_t\end{pmatrix}, \quad D_R^{-1}\mathcal{M}^d D_L = \begin{pmatrix}m_d & 0 & 0\\ 0 & m_s & 0\\ 0 & 0 & m_b\end{pmatrix}. \tag{3.37}$$

For quarks, a consequence of this diagonalisation is that it mixes the weak eigenstates and allows for flavour changing interactions within the electroweak sector with a current proportional to

$$\frac{g}{2}\begin{pmatrix}\bar{u} & \bar{c} & \bar{t}\end{pmatrix}_L \gamma_\mu V_{CKM} W^+_\mu \begin{pmatrix}d\\ s\\ b\end{pmatrix}_R + h.c. \tag{3.38}$$

where $V_{CKM}$ is the Cabibo-Kobayashi-Maskawa matrix which governs the strength of the flavour changing interactions [47].

$$|V_{CKM}| = |U_L^\dagger D_L| = \begin{pmatrix}|V_{ud}| & |V_{cd}| & |V_{td}|\\ |V_{us}| & |V_{cs}| & |V_{ts}|\\ |V_{ub}| & |V_{cb}| & |V_{tb}|\end{pmatrix} = \begin{pmatrix}0.974 & 0.225 & 0.004\\ 0.225 & 0.973 & 0.041\\ 0.009 & 0.041 & 0.999\end{pmatrix}. \tag{3.39}$$

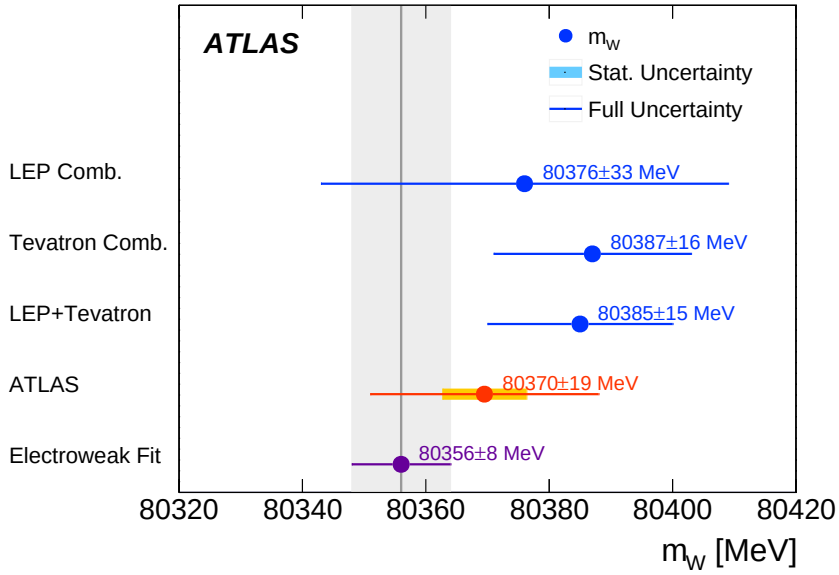



Figure 3.9: Current best measurements of $W$ mass including the global electroweak fit [48].

In the lepton sector of the Standard Model this rotation can be performed without introducing weak mixing due to the lack of neutrino masses. If Dirac neutrino masses are added, an identical procedure can be used to diagonalise the lepton masses at the expense of introducing neutrino oscillations. However, if the neutrinos are in fact Majorana fields, then a different procedure must be used. In either case, the result is a second mixing matrix (called the Pontecorvo-Maki-Nakagawa-Sakata or PMNS matrix) which shall not be explored further here.

### 3.3.4 The W and Z Bosons

Since a large part of this thesis will involve the $W$ and $Z$ bosons it is worth exploring their properties in more depth. The $W$ and $Z$ are spin-1 (vector) gauge bosons with masses of 80.4 GeV (see figure 3.9) and 91.2 GeV respectively [47]. Both $W$ and $Z$ decay to the full spectrum of fermions, with the sole exception of the top quark. Their branching ratios are given in table 3.3.

The W couples equally to all generations of leptons but for quarks $BR(W \to ij) \propto (3|V_{ij}|)^2$ (where $i \in \{u, c\}, j \in \{d, s, b\}$). As a result, decays to non-diagonal quark combinations are highly suppressed, leaving $W \to ud, cs$ as the primary hadronic decay modes. The $Z$ decays primarily to fermion-antifermion pairs; unlike the $W$ it

| $W$ Decay | Branching Ratio |
|---|---|
| $e\nu_e$ | $0.1071 \pm 0.0016$ |
| $\mu\nu_e$ | $0.1063 \pm 0.0015$ |
| $\tau\nu_e$ | $0.1138 \pm 0.0021$ |
| Hadrons | $0.6741 \pm 0.0025$ |

(a)

| $Z$ Decay | Branching Ratio |
|---|---|
| Invisible | $0.2000 \pm 0.0006$ |
| $\bar{e}e$ | $0.0363 \pm 0.00004$ |
| $\bar{\mu}\mu$ | $0.0366 \pm 0.00007$ |
| $\bar{\tau}\tau$ | $0.0366 \pm 0.00008$ |
| Up-type quarks | $0.116 \pm 0.006$ |
| Down-type quarks | $0.156 \pm 0.004$ |

(b)

Table 3.3: Primary decay channels for the W (a) and Z (b) bosons [47].

preserves flavour quantum numbers. The largest decay mode is to hadrons with an enhanced branching ratio to down-type rather than up-type quarks.

$$\frac{\frac{1}{3} \times BR(Z \to \bar{d}d + \bar{s}s + \bar{b}b)}{\frac{1}{2} \times BR(Z \to \bar{u}u + \bar{c}c)} = \frac{15.6}{11.6} = 1.34 \tag{3.40}$$

Single and pair produced $W/Z$ bosons have been observed at the LHC in run-2 and measurements of $\sigma_{tot} \times BR$ have been made for inclusive $W^{\pm}$ and $Z$ production [49–52]. The cross-section measurements made so far have most commonly been in the (relatively) less dominant lepton decay modes. Observing hadronic boson decay is typically more challenging due to the large multijet background. However, it was realised that hadronic decays were a useful tool in searches for high-mass particles [53] as the distinctive decay topology of boosted $W/Z$s could be used to discriminate against the multijet background. This has since become a common tool for exotic searches within ATLAS and CMS [54–57].

One additional property of the $W/Z$ bosons is that, as a result of absorbing the extra degree-of-freedom from the Goldstone bosons, the $W$ and $Z$ bosons may have both longitudinal and transverse polarisations. These states are primarily observed in colliders as differences in the angular distributions of the boson's decay products.

As the LHC is a proton-proton collider, the produced $W^{\pm}$ bosons are typically left-handed [58] leading to an asymmetry in the $W^{\pm} \to l^{\pm}\nu$ energy distributions. Boson polarisation is also a useful tool for discriminating between new models of physics which include new couplings to the $W$ and $Z$.

## 3.4 Beyond the Standard Model

Despite the rosy picture painted in the introduction to this chapter, we know that the Standard Model is not perfect. As an effective theory of the strong and electroweak forces, it has proven highly reliable, but even the obvious question of ''can we include gravity'' leads one down a road full of un-renormalizable infinities. There are also other outstanding issues in the theory including:

- What, if anything, cancels the large top contribution to the Higgs mass (the hierarchy problem)?
- Are there enough CP violating interactions to generate the observed matter-antimatter asymmetry?
- By what means do neutrino's gain mass?
- Why are there exactly three generations of fermions?

One of the primary goals of the LHC is to try and answer these questions (and others). In preparation for §7 we will give a brief overview of two ''beyond the Standard Model'' (BSM) theories. The Randall-Sundrum (RS) graviton and the Heavy Vector Triplet (HVT) models.

### 3.4.1 Extra Dimensions and the RS Graviton

A naive attempt to include gravity within the Standard Model leads to a non-renormalizable theory, i.e. one in which there are an infinite number of infinities which need to be cancelled. To (non-rigorously) see why this might be the case, consider the classical Einstein-Hilbert action

$$\mathcal{S}_{EH} = \frac{1}{2\kappa} \int d^4x R\sqrt{-g} \tag{3.41}$$

where $\kappa$ is Einstein's constant, $g = \det(g_{\mu\nu})$ and $R$ is the Ricci scalar (or Curvature scalar). In a full quantum theory of gravity, we would expect this to appear within

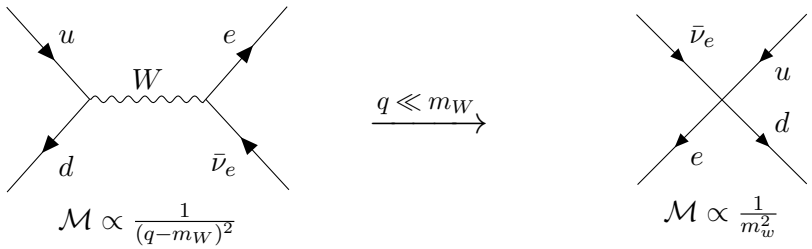



Figure 3.10: Beta decay in Fermi theory where we ''integrate out'' the $W$ propagator from full electroweak theory into a (non-renormalisable) contact interaction when $q \ll m_W$.

the Lagrangian similarly to the Maxwell action in QED with $\kappa^2$ playing the role of the coupling constant. However, in contrast to the dimensionless Standard Model couplings, $[R] = [m^2]$ and g is dimensionless so $[\kappa] = [m^2]$. Recall the definition of the Ricci scalar

$$R = 2g^{ab}\left(\Gamma^c_{a[b,c]} + \Gamma^d_{a[b}\Gamma^c_{c]d}\right) \tag{3.42}$$

where $\Gamma^a_{bc} = \frac{1}{2}(\partial_b g^a_c + \partial^a g_{bc} - \partial_c g^a_b)$ are the Christoffel symbols. If we expand eqn. (3.41) in terms of $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, where $h_{\mu\nu}$ is a small deviation that becomes the graviton field, then it has the index-suppressed form

$$S = \frac{1}{2\kappa^2}\int d^4x(\partial h\partial h + h\partial h\partial h + h^2\partial h\partial h + \dots) \tag{3.43}$$

Obviously, this means that pieces such as the triple graviton coupling $h\partial h\partial h/\kappa^2$ do not have mass-dimension 4, leading to non-renormalisable terms [59]. Hence, most quantum field theories of gravity are written as effective field theories, whose divergences are suppressed at experimentally accessible scales and are treated as the low-energy limit of some unknown theory. This approach is similar to the modern understanding of the Fermi theory of the weak interaction in §3.3 (figure 3.10). This method gives reasonable results provided that the energy scale of the interaction is much smaller than the mass scale of the high-energy theory.

The Randall-Sundrum graviton model [60] approaches a theory of quantum gravity in a less straight-forward fashion but in doing so they not only find an effective theory of gravity but also solve the hierarchy problem. Building on the idea behind the Kaluza-Klein model, they posit that the universe is in fact 5-dimensional, with the $5^{\text{th}}$ taking for the form of an $S_1/\mathbb{Z}_2$ orbifold (where $S_1$ is a circle). Further, they assume that at each stationary point on this orbifold ($x_4 = 0, L = \pi R$) there is a

3+1-dimensional brane. The metric for such a setup is given by

$$ds^2 = e^{-2k|x_4|}\eta_{\mu\nu}dx^\mu dx^\nu + d(x_4)^2 \tag{3.44}$$

A notable feature of this metric is that if we confine the Standard Model particles to the brane at $x_4 = L$ then there is a suppression of the Higgs vacuum expectation value (v.e.v.) of $\tilde{v} = e^{-2kL}v$. For chosen values of the curvature factor $k$ and brane separation $L$ this naturally solves the hierarchy problem (and equivalently, that of the scale difference between gravity and the other forces). The suppression factor leaves the $x_4 = 0$ brane at the Planck scale but reduces the $x_4 = L$ brane to the TeV scale. $k$ and $L$ are free parameters which can be chosen together such that $kL \approx 10^{16}$. For convenience, $k$ is usually written as $\tilde{k} = k\overline{M}_{Planck}$, where $\overline{M}_{Planck}$ is the reduced Planck mass, or simply given as $k/\overline{M}_{planck}$.

The experimental signature of such a model is a single massless spin-2 graviton and an infinite tower of massive excited gravitons beginning at the TeV scale. The mass of the lowest excited mode is

$$m_{Gr} = 3.83\tilde{k}e^{-kL}\overline{M}_{Planck} \sim \text{TeV} \tag{3.45}$$

The best current limits on such a particle come from the search for $G \to \gamma\gamma$ decays at CMS that exclude gravitons with a mass of 1.95 TeV (4.45 TeV) for $\tilde{k} = 0.01(0.2)$ [61].

If the Standard Model particles are allowed to propagate away from the brane (in the bulk), then these limits, alongside those from FCNC, can be avoided [62]. In this ‘‘bulk RS’’ model, the fermions are localised between the branes based on their observed masses, the 3rd generation close to the Planck brane and the 1st and 2nd generation fermions close to the TeV brane. Because of this, the graviton's decays to the light fermions are highly suppressed and its couplings to the Higgs, W and Z become large. This makes searches for exotic resonances decaying to vector bosons ideal channels to place limits on this process. The current best limits on the bulk RS model come from $G \to WW \to qql\nu$ decays at ATLAS, where $m_G > 1.75$ GeV is found (for $\tilde{k} = 1.0$) [63].

As an aside, there is a second, related, model referred to as RS2 [64] (the former being called RS1 in comparison), in which all particles, including the graviton, are localised around the Planck brane and the other is taken to infinity. While this loses the appealing feature of resolving the hierarchy problem, it is still an effective low-energy effective theory of gravity (assumed to be embedded within a full string

theory).

### 3.4.2 The Heavy Vector Triplet Effective Theory

Most experimental searches are primarily sensitive to the mass of a new resonance and its decay modes, rather than their spin or polarisation. The Heavy Vector Triplet (HVT) model [65] (and similar phenomenological models) exploits this by providing a simplified parameter set which on which limits can be placed. These phenomenological limits can then be reinterpreted in the context of more theoretically motivated models. The HVT model assumes an additional $SU(2)_L$ triplet of vector bosons, $V^{+,-,0}$, with zero hypercharge. These mix with the existing (pre-EWSB) $W$ triplet and couple to the Standard Model particles in a similar fashion. Post symmetry-breaking, we are left with the Standard Model and three additional massive bosons. The model contains the full set of CP-preserving dimension-4 operators and has nine free parameters, summarised in table 3.4.

| Parameter | Name |
|---|---|
| $g_V$ | Coupling strength |
| $m_V$ | Bare mass of the boson triplet |
| $c_H$ | Higgs coupling |
| $c_l$ | Lepton coupling |
| $c_q$ | Quark coupling (1st and 2nd generation only) |
| $c_3$ | 3rd generation quark coupling |
| $c_{VVV}$ | Trilinear self-coupling |
| $c_{VVHH}$ | Quadrilinear coupling to Higgs |
| $c_{VVW}$ | Trilinear coupling to (pre-EWSB) weak gauge fields |

Table 3.4: Free parameters of the HVT model.

The coupling strength may vary between $\sim 1 - 4\pi$ and each individual coupling is typically of order 1, except for $c_H$, which may vary more strongly than the rest. The case where $c_H \approx 1$ is known as the strong-coupling scenario and $c_H \ll 1$ as the weak-coupling scenario. $c_H$ is particularly important as it also controls the coupling to the Goldstone bosons and hence the branching ratio to the $W^{\pm}, Z$ bosons post-symmetry breaking (the $c_{VVW}/c_{VVHH}$ terms contribute minimally to searches as they require the creation of two of the new heavy bosons).

The masses of the new bosons are constrained by the gauge structure of the theory and the experimental measurements of the $W, Z$ masses. Taking $M_{\pm,0}$ to be the

masses of the new bosons post symmetry-breaking we have the two relations

$$\frac{m_{W,Z}}{m_V} \sim \frac{m_{W,Z}}{M_{\pm,0}} \lesssim 0.1, \quad \frac{M_\pm}{M_0} = |\cos\theta_W| \frac{m_Z}{m_W} = 1. \tag{3.46}$$

As the given parameters enter the Lagrangian pre-symmetry-breaking it is not immediately obvious how to relate these to experimentally measurable quantities, in practice experimental searches have set limits on two explicit models referred to as Model A and Model B that relate these parameters to measurable coupling strengths.

In Model A, an $SU(2)_1 \times SU(2)_2 \times U(1)_Y$ gauge group is broken to the Standard Model $SU(2)_L \times U(1)_Y$ by an additional doublet of scalar fields, with representation $(2,1)_{1/2}$ under the extended gauge group and v.e.v $f$. Working in a basis in which this symmetry breaking has occurred, it is found that:

$$\begin{gathered} g_V = g_2, \quad g_W^2 = \frac{1}{g_1^2} + \frac{1}{g_2^2}, \quad m_V = \frac{g_V f}{\sqrt{g_V^2 - g^2}}, \\ c_H = \frac{-g^2}{g_V \sqrt{g_V^2 - g^2}}, \quad c_F = \frac{-g_V}{\sqrt{g_V^2 - g^2}}. \end{gathered} \tag{3.47}$$

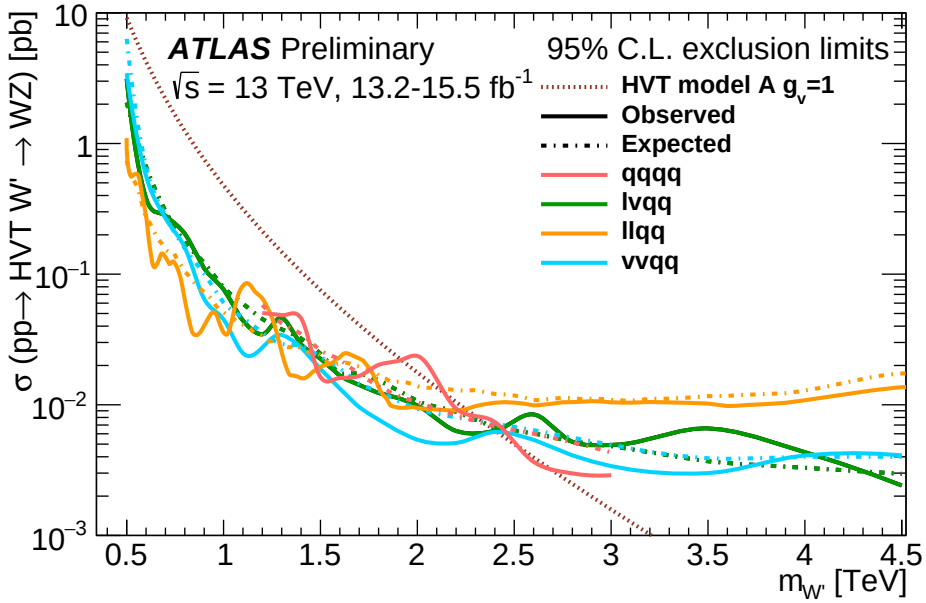
The vector boson in model A primarly decay to gluons with approximately 2% coupling to each of the Standard Model vector bosons and fermions.

Model B describes the lightest vector resonances found in a minimal composite Higgs model where the observed Higgs is a pseudo-Nambu-Goldstone boson emerging from an unknown strong force. In this scenario, a single spin-1 field $\rho$ is introduced, alongside the free parameter $f$ that controls its coupling to the light Higgs. Following calculation of its Lagrangian the terms can be matched with the parameters of the HVT model and in this case:

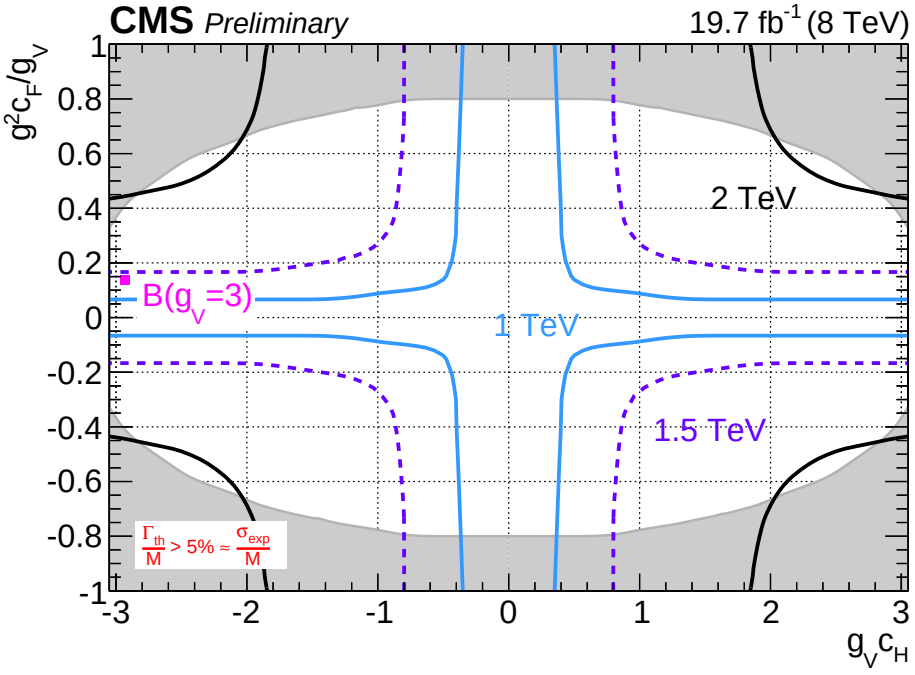
$$\begin{gathered} g_V = g_2, \quad g^2 = \frac{1}{g_1^2} + \frac{1}{g_2^2}, \quad m_V = m_\rho, \\ c_H = \frac{g_V}{\sqrt{g_V^2 - g^2}} \left[ \frac{g^2}{g_V^2} - \frac{m_V^2}{g_V^2 f^2} \right], \quad c_F = \frac{g_V}{\sqrt{g_V^2 - g^2}}. \end{gathered} \tag{3.48}$$

The Model B heavy vector triplet mainly decays to the Standard Model Higgs and vector bosons, as its decays to fermions are heavily suppressed.

Experimentally, limits are either set on these models directly or set on the generic terms $g^2 c_F / g_V$ and $g_V c_H$ and then reinterpreted in terms of the models shown above. Examples of both are shown in figure 3.11.




(a)




(b)

Figure 3.11: Example limits set on the HVT model. (a) shows the limits on Model A for a fixed $g_V$ by run-2 diboson searches at ATLAS [38] and (b) shows limits on the $g^2 c_F/g_V, g_V c_H$ parameter space derived from a run-1 search for $WH \to \ell\nu bb$ resonances at CMS [15].

# Chapter 4

# From Quarks to Jets

It was established in §3.2.2 that observing a bare quark is somewhat tricky. In fact, this is only the beginning of a physicist's problems, as shall be shown in §4.1. Even in the weakly-bound UV regime, ambiguities in quark final states due to the emission of soft or collinear gluons make the bare quark calculations problematic. The solution to this is to define a new type of physical observable - jets - which are both theoretically well-defined and experimentally accessible. This chapter will (briefly) review some of the theoretical aspects of quark emissions, including parton showers and hadronisation (based primarily on [66]), alongside jet construction and substructure. As much as is possible, this chapter aims to present a view of jets that is independent of the nature of their constituents. The algorithms and tools presented are regularly applied to jets constructed from calorimeter deposits, tracks, truth particles, and more. An in-depth discussion of jet construction in ATLAS analyses is left for §5.

## 4.1 Observing a Parton

The principle of colour confinement was discussed briefly in §3.2.2, but the physical consequences were not explored. In practice, confinement means that the only direct observations of quarks or gluons we can make involve colour-singlet bound states - hadrons. It is natural, therefore, to ask the question: what is the relationship between a parton which appears in the outgoing leg of a Feynman diagram and the hadrons observed?

The answer (as used in the simulation of this behaviour for LHC collisions) comes

from a similar factorisation argument as was used in §3.2.3 and is to treat the production, evolution and hadronisation of quark and gluon final states as independent processes. In a general purpose event generator such as Pythia [67], Herwig [68] or Sherpa [69] the generation process could be split into three main steps:

(1) Calculate (or retrieve) the matrix element for the desired hard interaction process;

(2) Evolve each parton from the interaction scale down to the hadronisation scale, typically $\sim$ 1GeV, emitting soft and collinear partons at each step, treat the probability of each emission as a Markov process (the parton shower);

(3) Combine the remaining partons into hadrons according to a hadronisation model. Since $\alpha_s \sim O(1)$ at this stage, the models are empirical rather than derived from first principles.

Of course, this is not the complete simulation process. The effects of initial state radiation and proton PDFs must be taken into account and the hadrons formed may be unstable and decay further. The complete process for a typical event is shown in figure 4.1. Additionally, the restriction to only soft and collinear emissions in step (2) seems to leave some phase space unaccounted for. In practice, non-soft/collinear emissions are treated as additional outgoing partons at the matrix element level and carefully merged to prevent double-counting.

### 4.1.1 Parton Showers and DGLAP

In the collinear limit, where the splitting angle $\theta$ between the child partons is much smaller than any separation in the hard process, the differential cross-section of any QCD process with a parton splitting can be factorised into a form proportional to the bare cross-section $\sigma_0$ and the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) splitting kernel $P(z, \phi)$:

$$d\sigma = \sigma_0 \frac{\alpha_s}{2\pi} \frac{d\theta^2}{\theta^2} dz P(z, \phi) \frac{d\phi}{2\pi} \tag{4.1}$$

where $\phi$ is the azimuthal angle of splitting around the parent parton and $z$ is the energy fraction of the leading child parton. The spin-averaged DGLAP equations

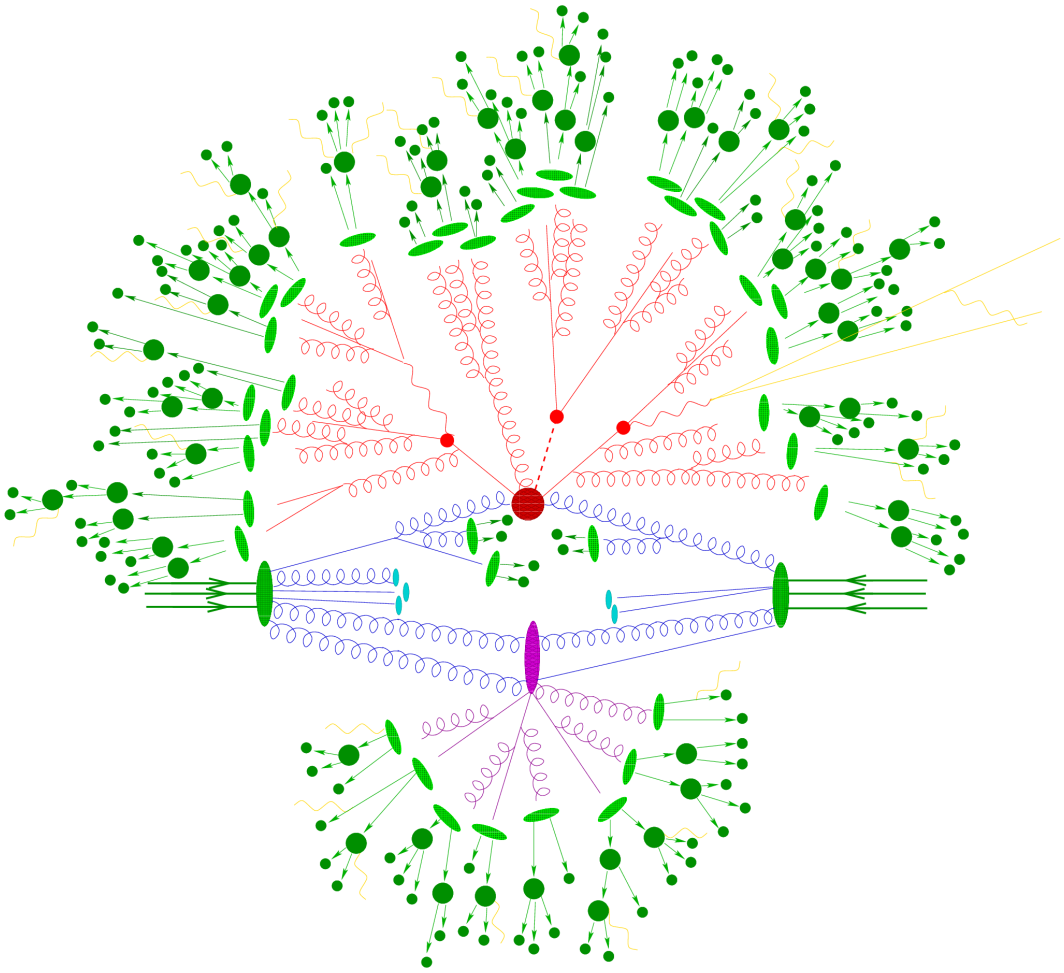
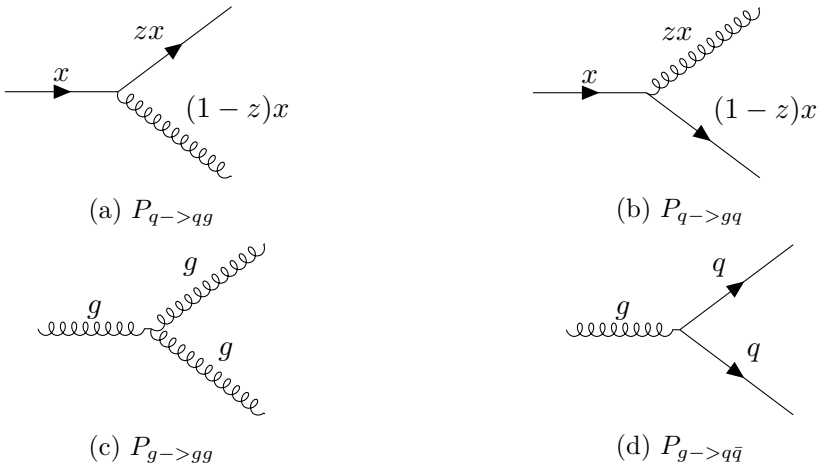


Figure 4.1: The structure of a typical (if low-multiplicity) MC event including initial state radiation and a secondary hard scatter. Particles from the hard collision (red blob) are showered and then hadronised, with unstable hadrons decayed to a stable final state. The underlying event contributions are also taken into account during the showering stage and contribute to the final event. Figure reproduced from [70].




(a) $P_{q->qg}$ (b) $P_{q->gq}$

(c) $P_{g->gg}$ (d) $P_{g->q\bar{q}}$

Figure 4.2: Feynman diagrams for the DGLAP splitting equations.

for massless quarks are

$$P_{q\to qg}(z) = C_F \frac{1+z^2}{1-z}, \qquad P_{q\to gq}(z) = C_F \frac{1+(1-z)^2}{z},$$
$$P_{g\to gg}(z) = C_A \frac{z^4+1+(1-z)^4}{z(1-z)}, \qquad P_{g\to q\bar{q}}(z) = T_R(z^2+(1-z)^2).$$

for the splittings shown in figure 4.2. Observe that $P_{q\to qg}$ is divergent in the limit $z \to 1$ and $P_{q\to qg}$ diverges as $z \to 0$, i.e. when the emitted gluon's momentum approaches zero. $P_{g\to gg}$ diverges in both of these cases. This can be resolved by recognising that the particles must be ''observed'' at some arbitrary resolution. Two partons with a separation below this limit would be unresolved and indistinguishable from the non-split case. The total probability of not observing an emission can be written as

$$P(\text{unresolved}|\text{split}) + P(\text{not split}) \tag{4.2}$$

and is finite when one-loop corrections to $P(\text{not split})$ are included. This relation is used to derive the Sudakov form factor which gives the probability of no observable splitting occurring between the scales $Q$ and $q$

$$\Delta(Q^2, q^2) = \exp\left[-\int_{Q^2}^{q^2} \frac{dk^2}{k^2} P(k^2)\right] \tag{4.3}$$

The consideration of whether partons can be physically resolved will be returned to in §4.2 in the context of jets.

The Sudakov factor is a core component of parton shower algorithms. Each parton

is evolved from the hard interaction scale downwards in a step-wise manner. For each step, the probability of a parton's non-emission is calculated independently of any other particles in the event or that parton's previous splitting history using the Sudakov form factor. If an emission is made then the child parton is also evolved similarly. Unfortunately, the ordering with which partons are evolved is non-trivial and must be carefully considered.

This is due to the divergence in the DGLAP equations when $z \to 0$, corresponding to the emission of low energy gluons referred to as soft emissions. This has a rather different form than the collinear splitting as the factorisation is only at the amplitude rather than the cross-section level. There are two common approaches for handling this. Angular-ordered showers (e.g. Herwig) begin from the DGLAP equations above and modify the splitting process to account for soft radiation correctly. Soft radiation emitted outside of a colour-connected cone formed from a parton splitting sees only the overall colour charge of that cone, or equivalently, it sees the colour charge of the parton before its splitting. By ordering the parton splittings by angle, the soft radiation is correctly accounted for.

A second approach (used by Sherpa) is to use a virtuality-ordered colour dipole model that, rather than using single $1 \to 2$ splittings, considers colour-anticolour pairs. Gluons are treated as colour-anticolour dipoles and have a single emission, quark-quark pairs emit a single gluon, and two quarks and a gluon are treated as two dipoles. Each emission splits the system into two further colour-anticolour dipoles. The emission of soft gluons from such a dipole is then given by

$$d\sigma \approx \sigma_0 C_A \frac{a_s}{2\pi} \frac{dk^2}{k_\perp^2} dy \tag{4.4}$$

where $y$ is the rapidity.

Both angular-ordered and colour dipole techniques are also used to generate ISR. In the case of an angular-ordered shower, this is done by reversing the splitting equations and solving to find an initial state that gave rise to the interaction process. In the colour dipole model there is no explicit ISR but the remnants of the incoming hadrons remain connected to the interacting quarks and radiate accordingly; this turns out to correctly account for the effect.

The parton shower algorithms discussed are a leading logarithmic approximation, but can include the all-orders resummation of loop-contributions to gluon emission by evaluating $\alpha_s$ as $\alpha_s(k_T)$. Showers in Pythia and Herwig are performed in the

limit $N_c = \infty$, errors due to which are approximately $1/N_c^2 \approx 10\%$. Note that electromagnetic contributions are not included (though they can be handled by external algorithms such as PHOTOS [71]) but this is also an $\frac{\alpha}{\alpha_s} \sim 10\%$ effect.

### 4.1.2 Hadronisation

Of course, the parton showering only takes the simulation to an energy around $1GeV$, where perturbation theory breaks down, and the partons form hadrons. Since a theoretical description of the process is beyond current techniques, a phenomenological model, fit to experimental data, must be used. The precise model varies from generator to generator (and often there are multiple options within generators). Two models commonly encountered are the Lund string model (Pythia) and the clustering model (Herwig), but many other models make different assumptions, such as Dipsy's rope model or the EPOS hydrodynamic model.

In addition to obtaining the correct description of jet-level variables such as the angular or $p_T$ distributions, hadronisation models must also produce a broad range of mesons and baryons at a rate which (hopefully) matches data. Typically many parameters are used to tune each model to fit, but due to the universality of the hadronisation stage, these can be obtained from a small number of measurements and applied widely.

## 4.2 Jet Algorithms

Jet algorithms are used to cluster the hundreds of final state hadrons into a much smaller number of jets, which more closely resemble the original particles. An ideal algorithm would exactly invert the parton shower and hadronisation process, be theoretically calculable, be unaffected by pileup and radiation from the underlying event, and have low computational cost. Sadly, such an algorithm has not yet been found and the ''optimal'' algorithm is strongly dependent on both the physics being examined and the environment in which the collisions take place. Common jet algorithms fall into two categories: sequential reclustering algorithms, which cluster particles one-by-one based on some ranking criteria, and cone algorithms, which cluster particles in geometrical regions around a set of axes (often using complex methods to find the ideal axes with physically sensible, stable results).

To be used in perturbative QCD calculations, the algorithm must be protected

against the collinear and soft singularities identified in §4.1.1. This requirement gives two safety criteria:

- Collinear Safety: Splitting an input into two collinear inputs must not change the output.
- Infrared (Soft) Safety: The addition of a soft ($p_T \approx 0$) input must not modify the output.

Algorithms meeting these criteria are described as IRC-safe. It should be noted that IRC-safety is not an absolute requirement; while lepton colliders typically use IRC-safe sequential reclustering algorithms such as the Durham algorithm, hadron collider experiments have historically used non-collinear-safe cone algorithms due to their lower computational cost (a naive implementation of sequential reclustering algorithms for $n$ particles requires $O(n^3)$ operations). Iterative algorithms are also guaranteed to make use of all particles in the event. For $ee$ collisions, which have no underlying event interactions like those of the spectator quarks in proton collisions, this seems to be a useful property.

IRC-safety is required for exact comparisons between experiment and theoretical calculations but the perturbative order at which this becomes a problem depends on the process being studied. Soyez [72] introduced two new algorithms, the SISCone (Seedless-Infrared-Safe-Cone) and the anti-$k_T$ algorithms as IRC-safe replacements for previous cone algorithms that preserved many of their useful properties, such as the well-defined, roughly conical, area. However, increases in computing power and the introduction of FastJet [73] in 2008 have reduced the computational complexity of sequential reclustering algorithms such that they are now practical even in high-multiplicity environments.

The primary jet algorithms used in ATLAS belong to the $k_T$ family. These are IRC-safe sequential reclustering algorithms controlled by two continuous parameters, $\alpha$, which controls the reclustering order and $R$, which controls the size of the jet. The algorithm is as follows: first calculate a common distance metric between each pair of inputs, known as pseudojets, and between each pseudojet and the beam ($B$)

$$d_{ij} = min(p_{T,i}^{2\alpha}, p_{T,j}^{2\alpha})\frac{\Delta_{i,j}}{R}, \tag{4.5}$$

$$d_{iB} = p_{T,i}^{2\alpha} \tag{4.6}$$

where $\Delta_{i,j}$ is the distance between the pseudojets in $\eta$-$\phi$ space. Then, pick the ‘‘closest’’ distance and if it is a pseudojet pair, merge them, otherwise, if it is $d_{iB}$,

call $i$ a jet and remove it from the clustering. The distances are then recalculated for the remaining inputs. This process repeats until no pseudojets are left.

There are three commonly used versions of this algorithm, corresponding to the choice of $\alpha = 1, 0, -1$ and known as the $k_T$, Cambridge-Aachen (C/A) and anti-$k_T$ algorithms respectively, and they show substantial differences in their behaviour. Since $\alpha$ is continuous it is possible to use non-integer values, but the resulting behaviour is qualitatively similar to the three primary cases.

The $k_T$ ($\alpha = 1$) variant preferentially clusters soft radiation before merging the remaining hard constituents, which, in the absence of additional contributions from pileup or the underlying event, mirrors the QCD radiative process. It is useful for accurately finding the underlying (sub)structure of the jet. Unfortunately, it is also highly susceptible to the addition of soft radiation and the jet area can fluctuate greatly as further radiation is added. In ATLAS it is primarily used with a very small $R$ parameter (typically 0.2) to recluster already identified jet constituents into resolved subjets.

The Cambridge-Aachen ($\alpha = 0$) algorithm is purely spatially dependent, effectively ordered by splitting angle, and has proven useful for resolving substructure within the jet while being more stable than $k_T$ when pileup or underlying event (UE) radiation (soft interactions between the constituents of the colliding protons not involved with the interaction of interest) is added. It has been found that the pseudojets found by sequentially undoing the clustering can be used to tag a jet as multi-prong W/Z or Top decay or a one prong quark/gluon jet, such as in the BDRS procedure [53].

The anti-$k_T$ ($\alpha = -1$) algorithm is the primary jet algorithm used by ATLAS analyses. It preferentially clusters hard radiation then picks up the soft radiation around the resulting axis. This gives a conical shape but with the harder jet taking a ‘‘bite’’ out of the softer when two jets would otherwise overlap. Jets constructed by the anti-$k_T$ algorithm are well-resolved and more resilient to pileup and underlying event than $k_T$ or C/A jets but their clustering order does not give useful information about the underlying QCD structure.

It is relatively easy to see why the clustering order of $k_T$ and C/A algorithms provide kinematic and spatial information about the jet substructure respectively. Consider only the final clustering step, for a $k_T$ jet the clustering depends on $\min(p_{T,i}, p_{T,j})\Delta_{ij}$ so the softest radiation will have been captured first and the final clustering is likely to be between two pseudojets with the $p_T$ split reasonably evenly between them,

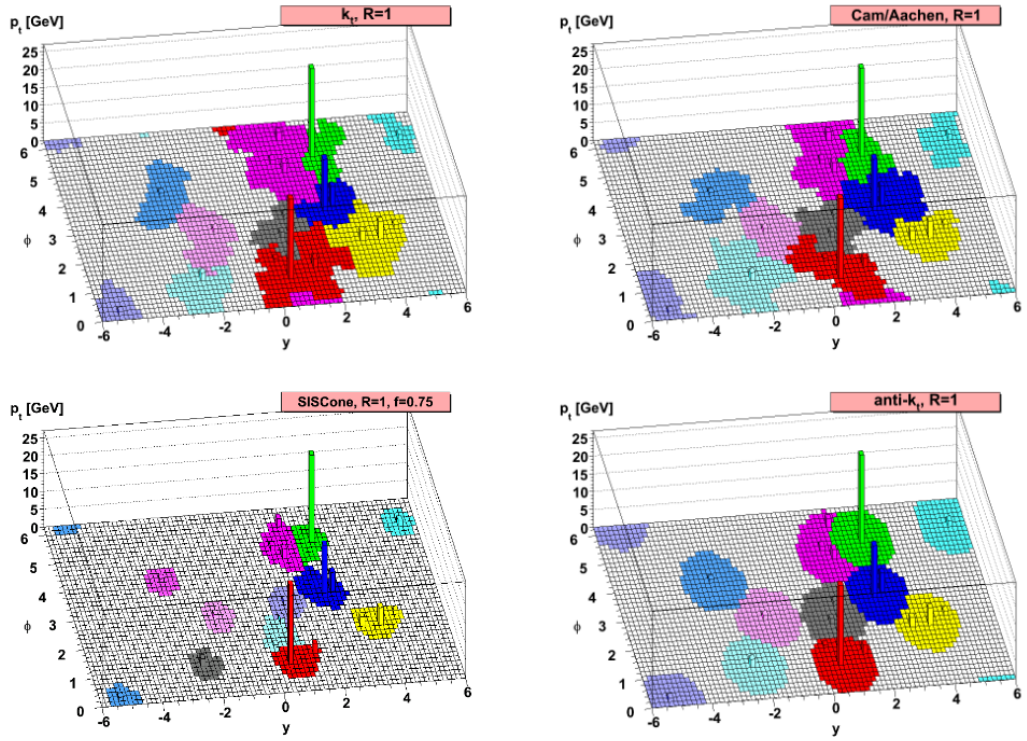



Figure 4.3: The jets constructed with the SISCone, $k_T$, C/A, and anti-$k_T$ algorithms with $R = 1$ in a single parton-level event with additional low-energy partons ("ghosts"). The catchment area for each jet is highlighted by colour. The area of the $k_T$ and C/A jets would be substantially different with a different set of ghosts whereas the anti-$k_T$ and SISCone algorithms produce more stable, conic, areas. Figure reproduced from [74].

closer to the original splitting of a hard parton or a W/Z decay. In contrast, for a C/A jet the final step would typically combine a high-$p_T$ central pseudojet and a (likely soft) pseudojet close to the jet radius but stepping back through earlier clusterings can find similar, $p_T$-balanced pseudojets in a two-prong decay.

The characteristic clustering behaviour of all of these algorithms (and SISCone) is shown graphically in figure 4.3. While discussion of it has been avoided so far the choice of ''R'' is significant for all jet algorithms and must find a balance between capturing the decay products of the hard process being studied while minimising contributions from pileup and the underlying event (UE), a term describing interactions that take place during a collision which are not the primary hard interaction. Jets with $R = 0.4, 0.5$ have better resolution when studying quark-initiated jets but larger $R = 0.8, 1.0, 1.2$ jets are used for capturing the decay of boosted bosons, tops and gluons [75]. The construction and calibration of jets within ATLAS will be discussed in §5.

## 4.3 Jet Substructure

As has been alluded to in the previous section, there are advantages to thinking of jets not just as singular objects but also making use of the distribution of their constituents. First, consider particles such as the W/Z, Top and Higgs (or other more exotic particles) that can decay to two or three quarks. If the originating particle is sufficiently ''boosted'', then its decay products will be produced in a narrow cone which could be identified as a single jet by the anti-$k_T$ algorithm. When searching for these objects, we can use the internal (sub) structure to ''tag'' jets with two or three prong-like structure and suppress the primarily one-prong QCD background. Secondly, we know that there will be pileup and UE contributions to the jet and that this can have a dramatic impact on physical observables. The jet mass, in particular, is very sensitive to the effects of soft radiation at a significant angle to the jet axis. ''Grooming'' algorithms have been designed to remove unwanted particles from jets.

Both substructure and grooming are only usually applied to ''large-R'' jets where $R > 0.8$, rather than the typical $R = 0.4/0.5$ (''small-R'') jets. Small-R jets are less useful for boosted searches due to a greater loss of the decay products outside of the jet cone, and their smaller size makes them inherently more resilient to pileup. This section will briefly cover grooming techniques and substructure variables, focusing

on the ones used later in this thesis.

### 4.3.1 Jet Grooming

Jet grooming algorithms take a jet constructed with one of the algorithms presented in §4.2 and try to remove all constituents which did not originate from the hard process. Doing so decreases the pileup's effect on the jet's reconstructed energy, momentum and mass. Several grooming algorithms have been investigated by both the CMS and ATLAS collaborations, but all use similar methods of reclustering a jet's constituents and keeping only those which pass some selection criteria. The grooming procedures are shown graphically in figure 4.4.

**Trimming**
Jet trimming [76] reclusters an input jet into smaller $k_T$ subjets and tries to remove the contributions of soft pileup and UE radiation by rejecting any subjets containing less than a fixed threshold of $p_T$. The algorithm parameters are the subjet radius $R_{sub} < R$, the dimensionless $p_T$ fraction, $f_{cut}$, and an associated hard scale $\lambda_{hard}$. The algorithm is:

1. Given a jet, $J$, obtain a list of its constituents $L$.
2. Recluster $L$ into subjets using the $k_T$ algorithm with a radius $R_{sub}$.
3. Reject any subjets with $p_{T,subjet} < f_{cut}\lambda_{hard}$.
4. Recluster the remaining subjet constituents using the original jet algorithm.

In ATLAS the $\lambda_{hard}$ is usually set to the $p_T$ of the jet.

**Pruning**
Jet pruning [77] considers the behaviour of a jet algorithm and cuts (''prunes'') away pseudojets using kinematic discriminants which prevent clusterings characteristic of the jet algorithm or QCD jets rather then heavy particle decays, specifically soft, wide-angled radiation. The algorithm has two parameters, $D_{cut}$ and $z_{cut}$, that control how aggressively it cuts in $\Delta R$, and $p_T$ space respectively. The algorithm is as follows:

1. Given a jet, $J$, obtain a list of its constituents $L$.
2. Run either the $k_T$ or C/A algorithm on $L$ and for each merging $ij \to p$ check $\frac{\min(p_{T,i},p_{T,j})}{p_{T,p}} < z_{cut}$ and $\Delta R_{ij} > D_{cut}$.

3. If either cut is passed veto the merging and discard the softer constituent. Continue with the jet finding procedure.

4. Stop once the resulting ‘‘pruned’’ jet passes the jet finding algorithm.

**Mass-Drop Filtering**

The jet mass-drop/filtering algorithm (sometimes known as split-filtering) was originally proposed in [53] as a combined grooming/tagging algorithm. It was formulated specifically to identify C/A jets with a hard two-prong structure (originally $H \to bb$) and has two separate steps, the ‘‘mass-drop’’ to determine the jet's structure, using the jet's clustering history as mentioned briefly in §4.2, followed by ‘‘filtering’’, which reclusters the jet constituents to remove pileup and UE contributions. There are two parameters used for the algorithm, the mass drop, $\mu_{frac}$,[1] and the energy balance $y_{frac}$. The mass-drop algorithm is then as follows:

1. Given a C/A jet, $J$, undo the last step of the clustering to get the subjets $j_1, j_2$ where $m_{j_1} > m_{j_2}$.

2. Require a significant difference in mass after the splitting, $\frac{m_{j_1}}{m_J} < \mu_{frac}$, and that the energy is evenly split between the subjets, $\frac{\min(p_{T,j_1}, p_{T,j_2})}{m_J^2} \times \Delta R_{12}^2 > y_{cut}$.

3. If both cuts are passed then consider $J$ to be the neighbourhood of the heavy particle splitting and continue to filtering. Otherwise set $J = j_1$ and repeat. If $j_1$ cannot be split further, reject the jet.

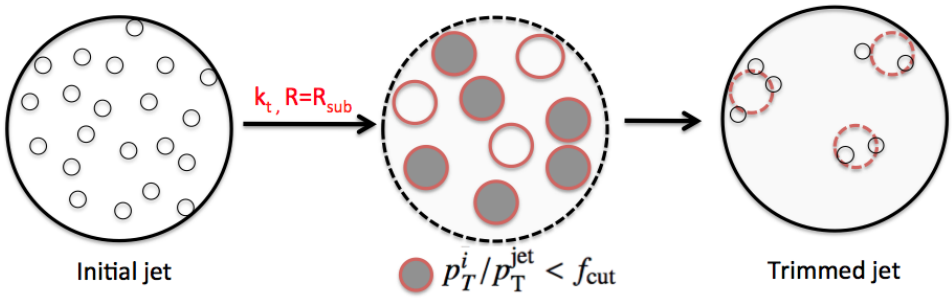
Once the candidate jet $J$ has been identified, its constituents are reclustered using C/A with $R = \min(0.3, \Delta R_{12}/2)$. The three hardest subjets are kept and combined into the final jet.

The first stage of this process has since been generalised, via the modified Mass Drop (mMDT) algorithm [79], to the Soft Drop [80] algorithm which can act as both a tagger and groomer.

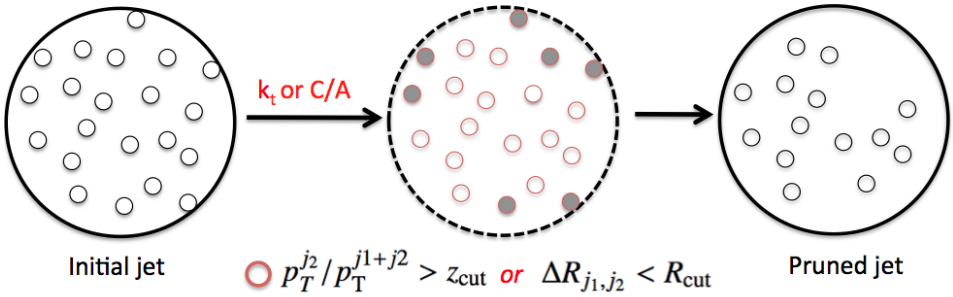
### 4.3.2 Substructure Variables

The internal, or sub, structure of a jet will inevitably be impacted by the nature of the particle(s) whose decay it contains. An obvious example is the jet mass, defined

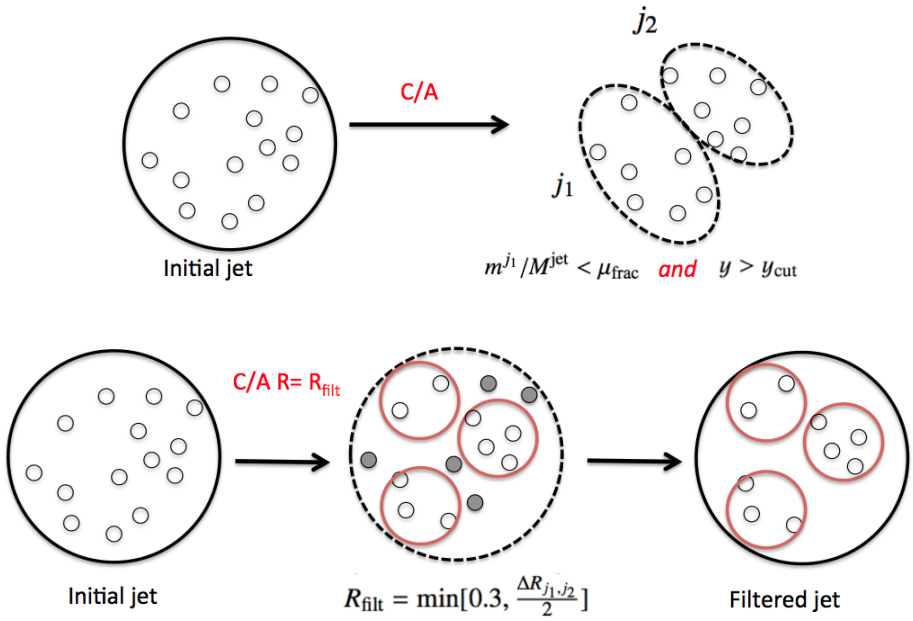
[1] It is perhaps ironic that, despite giving the tagger its name, it has been found that the optimal value for the mass-drop parameter is $\mu_{frac} = 1$, i.e. no mass-drop criteria should be applied [78]. As a result, modern generalisations such as $mMDT$ and soft-drop do not implement this cut.




(a)




(b)




(c)

Figure 4.4: Graphical examples of constituent removal by the presented jet grooming algorithms (a) jet trimming, (b) jet pruning, and (c) the split-filtering algorithm. Note that (c) is split between two phases, the mass-drop and the filtering. Typically only the second stage is used. Figures reproduced from [78]

in terms of the jet constituents, $j$, as

$$m_J^2 = \left(\sum_{j\in J} E_j\right)^2 - \left(\sum_{j\in J} \vec{p}_j\right)^2 \tag{4.7}$$

which is expected to be roughly $m_{particle}$ for heavy particle decays and approximately $0.2\sqrt{p_T}R$ [81] for QCD jets, making it a useful discriminant. However, additional non-associated radiation spoils this relationship by smearing the mass distributions of both heavy particle and QCD jets. This effect is particularly harmful for W/Z identification as the QCD mass peak shifts upwards into a similar mass range. Thankfully, the grooming procedures remove a large amount of this radiation, shifting the QCD mass range back down and sharpening heavy particle peaks.[2]

Beyond the jet mass, there is a significant amount of additional information stored within a jet's substructure. A gluon jet will contain more radiation than a quark jet due to the gluon's larger colour factor; Ws, Zs and Higgs will have a two-prong structure, but only Zs and Higgs jets are likely to contain b-quarks, a top decay will have a three-prong structure, and so on. There is now a substantial library of substructure variables which provide discriminating power between different types of jets. Here we will cover a few that feature in the rest of this thesis.

**Track Multiplicity**

The number of charged tracks, $n_{trk}$, within a jet can be used as a discriminant between quark, gluon and heavy particle jets. The gluon's enhanced colour factor means that we expect a greater emission rate compared to quark jets. This would be observed in the detector as additional charged tracks (uncharged emissions being missed by the tracking systems). Further, the colour connection between the two quarks in a W/Z/Higgs decay means that emission outside of the decay cone is suppressed and the total particle emission is lower than for single quark decay.

In ATLAS, this is measured by counting each jet's ghost-associated[3] tracks. In practice, $n_{trk}^{ungroomed}$ is more useful than $n_{trk}^{groomed}$ as cutting away the soft radiation associated with pileup and the UE also removes radiation associated with the true jet, smearing the $n_{trk}$ distribution.

[2] It is not necessarily the case that the groomed jet substructure is more useful than that of the ungroomed jet. Some phenomenological studies have suggested that the N-subjettiness (see §4.3.2) ratio of $\tau_{N,groomed}/\tau_{N-1,ungroomed}$ is more powerful than its groomed-only equivalent [82].

[3] Ghost-association, an alternative to clustering by $\Delta R$, is performed by adding the tracks into the clustering process with their $p_T$ set to 0. This clusters tracks within the jet without altering the jets found.

**N-subjettiness**

N-subjettiness [83] provides a measure to estimate the number of distinct subjets in a jet, based on the event level observable N-jettiness [84]. The basic procedure for calculating the set of N-subjettiness variables $\{\tau_1, \ldots, \tau_N\}$ is as follows. First, find a set of subjets $1, \ldots, N$ with $R_{sub} \ll R$ then calculate $\tau_N$ as

$$\tau_N = \frac{1}{\sum_{k \in J} p_{T,k} R} \sum_{k \in J} p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\} \tag{4.8}$$

where $k$ runs over the original jet's constituents and $\Delta R_{n,k}$ is the distance between $k$ and the axis of subjet $n$. $\tau_N$ will be small for any jet with $\leq N$ subjets but large otherwise. As such, the ratio $\tau_N/\tau_{N-1}$ gives a measure of how ''$N$-subjetty'' a jet is.

A key consideration for N-subjettiness is the procedure for finding candidate subjets. An exhaustive search minimising $\tau_N$ would give ideal results, but this was originally believed to be too computationally intensive. The initial paper suggests the exclusive $k_T$ [85] algorithm forced to return $N$ subjets, but it was later identified that the broadening axes (the set of axes which minimise the scalar sum of the jet momentum) [86] are the optimal axis choice. Further studies [87] have proposed the ''winner-takes-all'' ($\tau_{X,wta}$) axis which approximates the broadening axis but is also insensitive to recoil. The N-subjettiness variables was used by ATLAS for boosted top tagging in run-2 data [88] and investigated in for the purpose of W-tagging [89] as shown in figure 4.5.

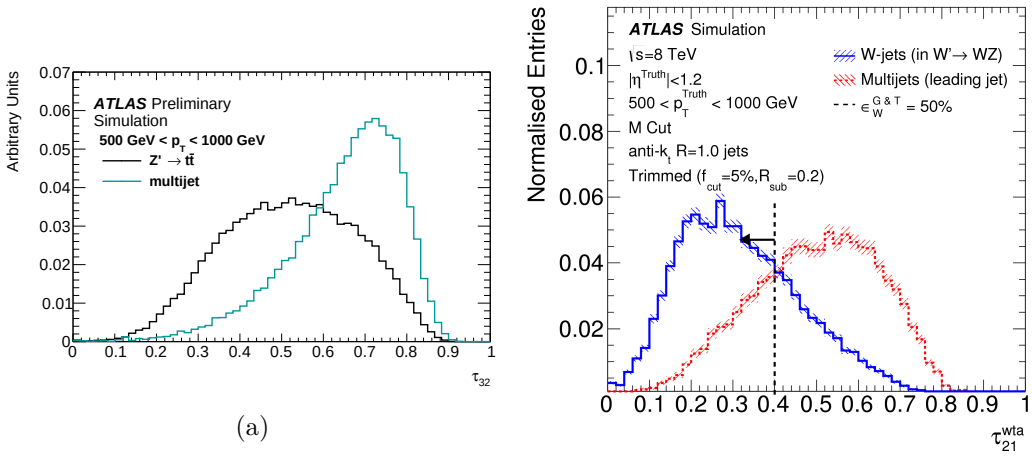
Further theoretical research on these variables has also suggested that the so-called ''dichroic'' $\tau_{21}$ ratio, where $\tau_2$ is taken from the ungroomed jet and $\tau_1$ is taken from the groomed jet, emphasising the colour radiation pattern and hard substructure respectively, may provide additional discriminatory power [82].

**Energy Correlation Functions**

Energy correlation functions [90] can be used to quantify the n-prong substructure in a similar way to N-subjettiness but without requiring the definition of subjet axes. The N-point energy correlation function (ECF) is defined as

$$ECF(N, \beta) = \sum_{i_1 < i_2 < \cdots < i_N \in J} \left( \prod_{a=1}^{N} p_{T,i_a} \right) \left( \prod_{b=1}^{N-1} \prod_{c=b+1}^{N} \Delta R_{i_b i_c} \right)^{\beta}. \tag{4.9}$$

If a jet $J$ has a $N+1$ prong structure then $ECF(N+1, \beta)$ should be substantially smaller than $ECF(N, \beta)$ as the $\Delta R$ terms will be small. Hence, the ratio $ECF(N+$




(a) (b)

Figure 4.5: Examples of the discrimination power provided by N-subjettiness between QCD, W and top jets with $500 < p_T^{truth} < 1000$. (a) shows the $\tau_{32}$ distribution for top and QCD jets, from [88], and (b) the $\tau_{21}^{wta}$ distribution for W and QCD jets, taken from [89].

$1, \beta)/ECF(N, \beta)$ behaves very similarly to the $\tau_N$ variables. This suggests a similar observable for tagging $N$-pronged jets:

$$C_N^\beta = \left(\frac{ECF(N+1,\beta)}{ECF(N,\beta)}\right) \times \left(\frac{ECF(N,\beta)}{ECF(N-1,\beta)}\right)^{-1} \tag{4.10}$$

An additional ECF-based variable was proposed in [91] which provides optimised discriminatory power between 1 and 2-prong jets.

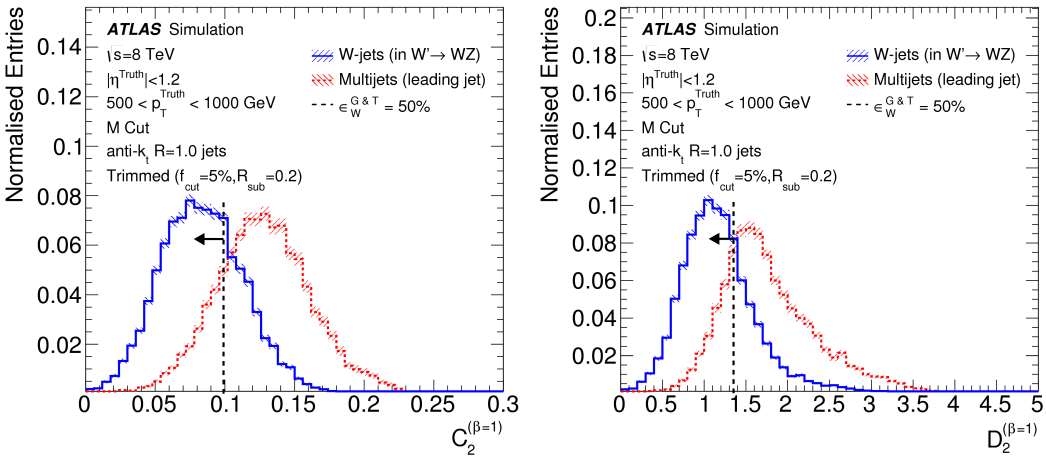
$$D_2^\beta = \frac{ECF(3,\beta)ECF(1,\beta)^2}{ECF(2,\beta)^3} \tag{4.11}$$

The distributions for $C_2$ and $D_2$ are shown in figure 4.6. For ATLAS searches using the $D_2$ variable, $\beta = 1$ has so far been found to be optimal [89, 92]. $\beta = 1$ is used exclusively throughout this thesis so this superscript will be dropped for the remainder of the text.

### Jet Width

The jet width is a dimensionless substructure variable that is highly correlated with the jet mass [93]. It is defined as

$$W = \frac{\Sigma_i \Delta R^i p_T^i}{\Sigma_i p_T^i} \tag{4.12}$$




(a) (b)

Figure 4.6: Examples of the discrimination power for trimmed, anti-$k_T$, $R = 1.0$, jets provided by the $C_2$ (a) and $D_2$ (b) variables for QCD and W jets with $500 < p_T^{truth} < 1000$ [89]. The distributions are similar for other jet algorithms.

where $\Delta R^i$ is the radial distance between each jet constituent and the jet axis.

### Angularity

Angularities [94] are a family of IR-safe (for $a < 2$) observables which describe the degree of symmetry of a jet's constituents. It is defined with a free parameter $-\infty < a < 2$, as

$$A_a = \frac{1}{M} \sum_i E_i \sin^a \theta_i [1 - \cos \theta_i]^{1-a} \tag{4.13}$$

$a$ controls whether quantity emphasises the constituents near the core ($a > 0$) or edges ($a < 0$) of the jet.

# Chapter 5

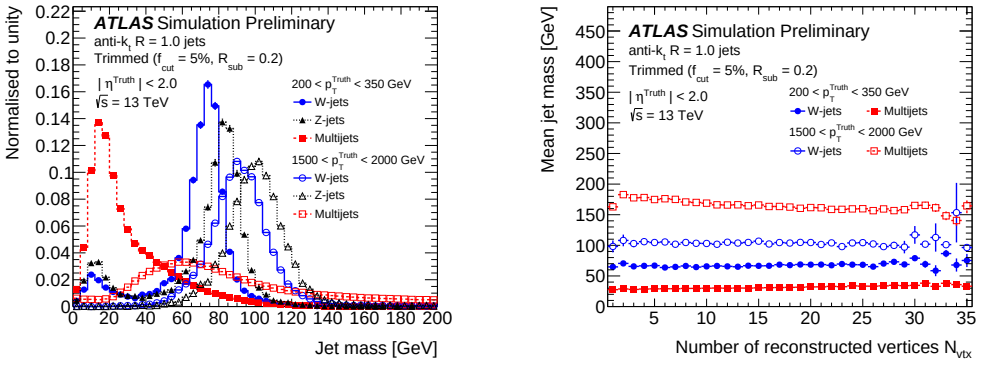
# Jet Reconstruction and Calibration

Large-R jets are primarily used within ATLAS to study hadronically decaying heavy particles, in particular, the W, Z, and Higgs bosons, and the top quark. Effectiveness in this role has a few essential requirements. First, the jet radius must be large enough to capture the decay products, and secondly, it must be possible to derive effective taggers for the desired initial particle. In practice, this second requirement demands good mass resolution and substructure information.

While the jet grooming algorithms discussed in §4.3.1 play an important role in making jets usable in the high-density environment of the LHC, they are not the whole story. Detector noise, pileup, the showering of different types of particle, and other effects such as dead material, alter the behaviour of jets constructed from raw calorimeter clusters. Some of this is corrected before reconstruction by combining the calorimeter clusters into calibrated ‘‘topological clusters’’ and, after construction, the jet energy and mass can be calibrated based on known detector behaviour. This chapter will briefly review jet reconstruction at ATLAS, but will primarily focus on a new approach to the second topic, jet calibration using (multivariate) neural networks.

## 5.1 Reconstruction

The standard large-R jet collections used within ATLAS are constructed from calibrated clusters of calorimeter cells called topoclusters [95]. A study in [92] found




(a) (b)

Figure 5.1: Figure (a) shows the measured calorimeter jet mass after trimming for truth W, Z, and QCD jets in two different bins of $p_T$ before calibration is applied. Figure (b) shows the mean jet mass of trimmed W and QCD jets as a function of the number of reconstructed vertices in the event [92].

that $R = 1.0$, anti-$k_t$ jets groomed using the trimming algorithm with $R_{sub} = 0.2$ and $f_{frac} = 0.05$ was the best choice of jet algorithm and grooming procedure for W/Z tagging under the pileup conditions expected during run-2. The performance of the trimming algorithm is demonstrated in figure 5.1. Taggers for Higgs and top jets were derived based on this optimised jet collection [88, 96]. Additionally, so-called ‘‘track’’ jets are also used, which take the collection of tracks matched to the primary vertex with $p_T > 400$ MeV as inputs. These are constructed using the same algorithms as the calorimeter jets.

### 5.1.1 Jet Calibration By ‘‘Numerical Inversion’’

Calibrating a jet requires some definition of the correct, or ‘‘true’’, value of the quantity being calibrated. At ATLAS this is typically defined in MC as the value when calculated using a jet constructed from particle-level constituents, prior to the detector simulation and reconstruction. Since this cannot be measured directly in data, the mapping between truth and reconstructed quantities is measured in MC and then validated and applied in data.

In ATLAS analyses, a calibration is applied to correct the measured jet energy and mass back to this ‘‘truth’’ scale and the mappings for both quantities are derived using the numerical inversion technique [97–99] on a simulated jet sample as follows.

Given a variable's response $\mathcal{R}_x = x^{reco}/x^{truth}$, the data is split into bins of $p_T^{truth}$, $|\eta^{reco}|$,

and $x^{truth}$ then $\langle R_x \rangle$ is extracted as the mean of a gaussian fit in each bin. The calibrated variable is defined as

$$x^{calib} = \frac{x^{reco}}{\langle \mathcal{R}_x \rangle (f^{-1}(x^{reco}))} \equiv c_x(E^{reco} \text{ or } p_T^{reco}, \eta^{reco}, x^{reco}) \cdot x^{reco} \tag{5.1}$$

where $f(x) = \langle \mathcal{R}_x \rangle(x) \cdot x$, the average reconstructed quantity given the truth-level quantity $x$, is the calibration function. $c_x$ is defined as the correction factor. This correction factor is derived separately in each bin of $|\eta^{reco}|$ by fitting the obtained means as a function of $p_T$ and $x^{reco}$ using a 2d gaussian kernel, as demonstrated in [100].

For the jet energy scale (JES), this is done directly on $E^{reco}$. For the jet mass scale (JMS), an identical procedure is used except the jets are binned in $(p_T^{truth}, |\eta^{reco}|, m^{truth})$ and $f \equiv f(p_T^{reco}, m^{reco})$ is fit to $\langle R_m \rangle$, $\langle p_T^{reco} \rangle$, and $\langle m^{reco} \rangle$ in each bin of $|\eta^{reco}|$. These calibrations are applied sequentially so the fully calibrated $p_T$ and mass are given by:

$$p_T^{reco} \to c^{JES}(E^{reco}, |\eta^{reco}|) \cdot p_T^{reco}, \tag{5.2}$$

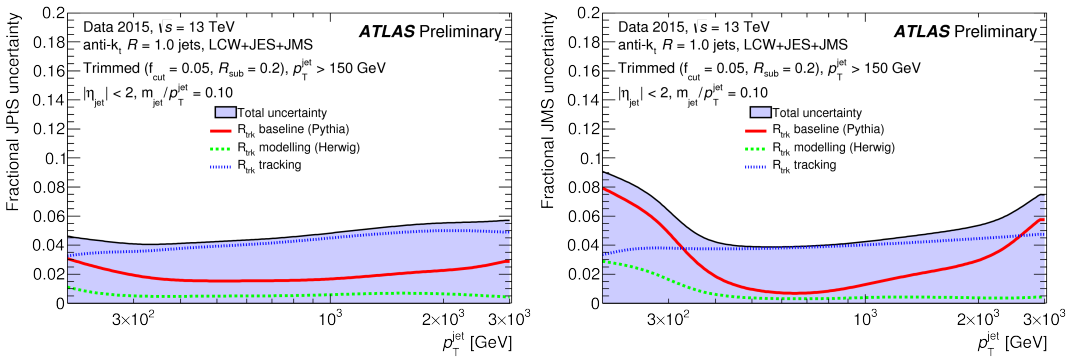
$$m^{reco} \to c^{JMS}(p_T^{reco}, |\eta^{reco}|, m^{reco}) \cdot c^{JES}(E^{reco}, |\eta^{reco}|) \cdot m^{reco}, \tag{5.3}$$

The jet energy is recalculated after the calibrations have been applied.[1] The systematic uncertainties on these calibrations, including on the extrapolation between data and MC, are derived in-situ and applied as systematic effects in any analysis using these calibrations. These are shown in figure 5.2.

### 5.1.2 Combined Mass

It was observed early in run-2 that the mass resolution of large-R jets deteriorated dramatically at high $p_T$, typically above 1.5 TeV. This effect is now understood to be the result of the delta-R separation of jet constituents falling below the calorimeter resolution. Since the mass measurement depends strongly on the angular separation of the jet's constituents, the merging of multiple constituents into a single calorimeter cell can have a significant impact on the measured mass. The jet mass response (defined as $\mathcal{R} = m^{reco}/m^{truth}$), used for calibrating the jet mass, has a distinct ''double-peak'' shape in the affected region and the calibration using numerical

[1] The JES correction is derived in terms of energy simply because the calorimeter responds to energy rather than $p_T$, despite this, $p_T$ is more commonly used by analyses so that is treated as the primary variable during calibration.




(a) (b)

Figure 5.2: The $p_T$ (a) and mass (b) uncertainties for the calibrated large-R jets estimated in-situ using the track double ratio method as a function of the jet $p_T$. Uncertainties shown are for jets with $p_T > 150$ and $m/p_T = 0.1$ [101].

inversion is not well behaved as shown in figure 5.3.

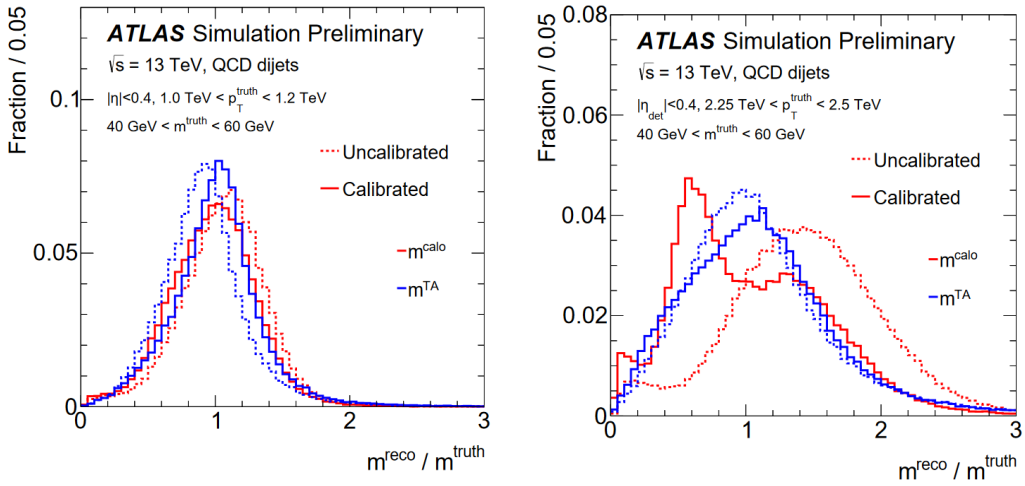
Since the problem is fundamentally caused by the angular resolution of the calorimeter, it was realised that the performance could be improved by using the significantly better angular resolution of tracks built using inner detector hits [98]. The ‘‘track mass’’, $m^{track}$ of a jet is calculated using the four-vector sum of all charged tracks associated with a jet (with the mass of each track set to $m_\pi$), but this does not account for neutral contributions to the jet. This is estimated from the ratio $p_{T,jet}^{calo}/\sum p_T^{track}$ and the ‘‘track-assisted mass’’ of the jet is defined as $m^{TA} = m^{track} \times p_T^{jet}/\sum p_T^{track}$. As shown in figures 5.4 and 5.3, the track-assisted mass resolution is better than that of the calorimeter mass at high $p_T$, and it does not suffer from the non-closure of the calibrated mass response. However, its resolution is worse at low $p_T$.

Good mass resolution and closure is achieved over the entire jet $p_T$ range by combining both calorimeter and track-assisted mass into the (imaginatively named) ‘‘combined’’ mass, defined as

$$m^{comb} = a \times m^{calo} + b \times m^{TA}, \tag{5.4}$$

$$a = \frac{\sigma_{calo}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}}, \quad b = \frac{\sigma_{TA}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}} \tag{5.5}$$

where $\sigma_{calo}^{-2}$ and $\sigma_{TA}^{-2}$ are the calorimeter and track-assisted mass resolution functions respectively. Since the inputs are calibrated, fixing $a + b = 1$ ensures that the combined mass is also calibrated, and the two constants smoothly interpolate between the region where the calorimeter resolution is superior and where the track-assisted




(a) (b)

Figure 5.3: The response for the calorimeter and track-associated jet mass for $1.0 < p_T^{jet} < 1.2$ TeV (a) and and $2.25 < p_T^{jet} < 2.5$ TeV (b) before and after calibration, reproduced from [98]. The high $p_T$ plot shows the clear non-closure of the jet calorimeter mass calibration at high $p_T$.

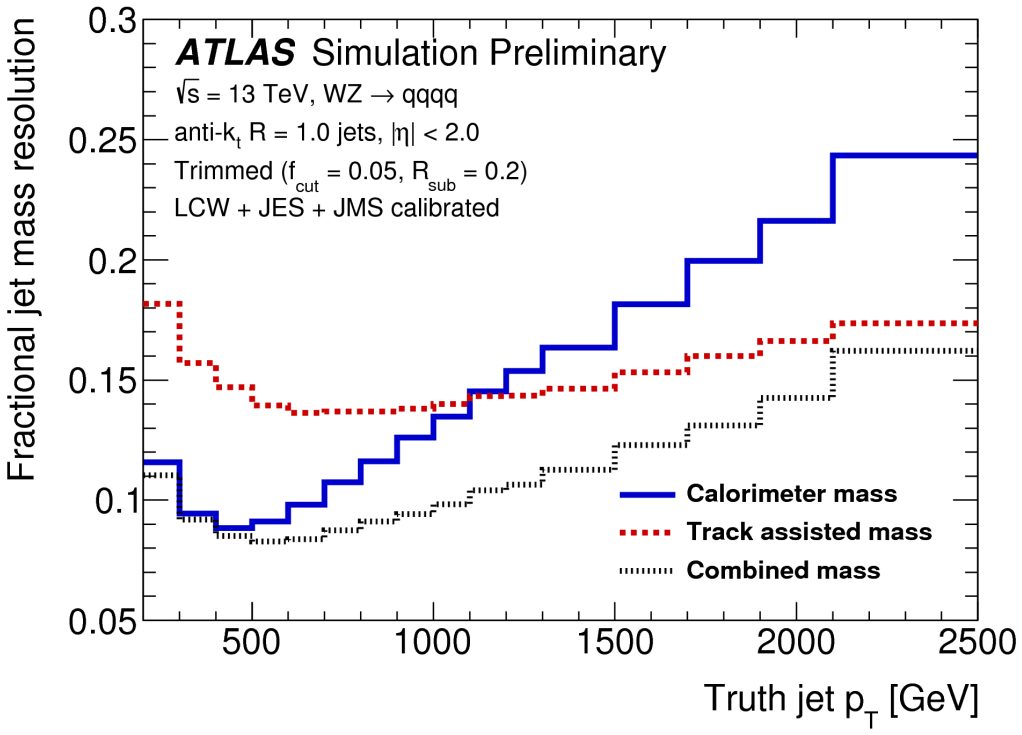



Figure 5.4: The fractional jet mass resolution for the calorimeter, track-assisted, and combined mass after calibration. Reproduced from [98].

resolution is better. The final resolution of the combined mass is shown in figure 5.4.

## 5.2 Multivariate Calibration

The current large-R jet calibration works well but it fails to take advantage of the additional information provided by the substructure of each jet. The pattern of energy deposits and hence calorimeter response may vary wildly between, for example, a two-prong W-tagged jet or a more evenly distributed, single-pronged QCD jet. Variables sensitive to the angular distribution of energy, mass being the primary example, could potentially benefit from using tracking information with superior angular resolution (as has already been seen with the combined mass). The numerical inversion technique outlined in §5.1.1 does not generalise well to more variables as each additional variable significantly increases the number of bins, and hence the required number of MC events. This section presents a proof-of-concept study into a different approach; rather than dividing the jets into many bins and fitting a simple response function in each, a neural network trained on substructure variables is used to approximate the true detector response function, across all $p_T$ and $\eta$ regions.

Neural networks have a history within particle physics going back to at least the early 1990s [102–104], primarily as classifiers. However, until recently they had fallen out of favour in comparison to boosted decision trees. Following the realisation that deep neural networks (DNNs) and convolutional neural networks (CNNs) outperformed other forms of machine learning in many situations, they have begun to make a resurgence in particle physics, particularly in the context of jet tagging [105, 106] but also in less traditional ways, for example calculating parton shower uncertainties [107], modelling NNLO corrections [108], and pileup mitigation [109]. Naively, the universal approximation theorem [110] suggests that any function can be approximated by a single layer neural network but in practice, deep neural networks are often able to approximate more complex functions than a single-layer network with the same total number of neurons.

The relative power of deep (defined as 3 or more layers) neural networks can be understood through two concepts. Firstly, non-linearities are substantially easier to approximate using a multilayered network than with a single layer network; in a deep network, a small change in the initial layer can lead to a strong response in

subsequent layers. Secondly, multilayer networks have a built-in concept of scale-dependence where each layer examines the input at a different ''scale''.[2] Since this study considers substructure and 4-vector inputs rather than raw spatial data it is the first of these properties that seems most useful.

For a single jet, the jet's response is defined for a target variable, $t$, as

$$\mathcal{R}_t = \frac{t_{reco}}{t_{truth}}. \tag{5.6}$$

We make the (seemingly reasonable) assumption that the calorimeter's response to a jet has a strong dependency on the pattern and location of the jet's energy depositions within the calorimeter, and that this dependency would manifest as a relationship between $\mathcal{R}_t$ and a jet's 4-vector information and substructure variables. There would undoubtedly be some variance that cannot be captured this way, either due to detector noise or deterministic behaviour not captured by a jet's substructure, but this study assumes that given a vector of input variables $\vec{v}$ it is possible to find an approximate function $O(\vec{v})$ such that

$$O(\vec{v}) \approx \langle \mathcal{R}_t \rangle (\vec{v}), \tag{5.7}$$

which may be used to calibrate a variable

$$t_{calib} = \frac{t_{reco}}{O(\vec{v})} \approx \frac{t_{reco}}{\langle \mathcal{R}_t \rangle}. \tag{5.8}$$

In essence this is simply a restatement of the general idea of calibration but I believe it is a useful form when considering machine learning calibrations. For calibration of a jet for target variable, $t$, and generic vector of input variables, $\vec{v}$, try find the function $O(\vec{v})$ by training a neural network on a MC dataset for which we already know the correct jet response. Note that conceptually this procedure is much simpler than the existing calibration.

This study will demonstrate several instances of this calibration procedure for jet energy and mass. It will also address measures of calibration quality, network architecture, the systematic selection of variables, and the impact of the MC generator on the final calibration. Since this is not intended for use in an actual analysis, the procedure used is further simplified from that given in eqn. (5.2) as only single step calibrations where energy and mass are calibrated separately are considered, though

[2] In fact, the scale dependence of DNNs can be understood as an application of renormalisation groups [111, 112]. It would be interesting to examine whether this is a useful property for the study of QCD more generally (for instance scale-dependent jet clustering algorithms).

in practice a combined or sequential calibrations would be simple extensions of this method.

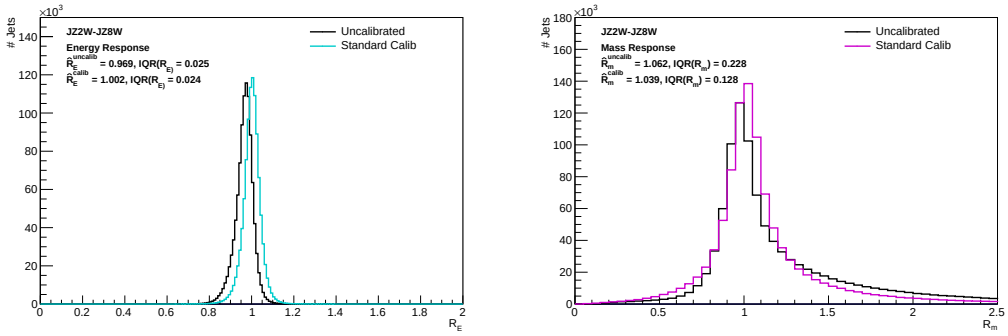
### 5.2.1 Calibration Quality

When talking about ‘‘improving’’ a calibration, it is necessary to define the measure of quality with which one is judging. Conceptually this is quite simple - we wish for the calibrated jet properties to be as close to the truth values as is possible - but there are several ways to define this. The current procedure in ATLAS is to evaluate large-R jet calibrations on two measures derived from the distribution of $\mathcal{R}_E$ and $\mathcal{R}_m$, the distributions of which are shown before and after calibration in figure 5.5. The calibrated jet ‘‘scale’’, taken to be the difference between the peak of a gaussian fitted to the core of the response distribution, and the calibrated jet ‘‘resolution’’, taken to be the standard deviation of the same fitted gaussian. This is a reasonable measure for $\mathcal{R}_E$ which is mostly symmetrical but $\mathcal{R}_m$ typically has a large upper tail. As such I prefer to use the following measures:

- Scale, $\hat{\mathcal{R}}$: the median of the response distribution after calibration.
- Resolution, $IQR(\mathcal{R})$: the interquartile range of the response distribution after calibration divided by two.
- Spread, $IDR(\mathcal{R})$: a new measure based on the interdecile range (similar to the interquartile range but the difference between the first and ninth deciles of the dataset) to give a more complete picture of the calibration resolution. It is defined as half the interdecile range of the response.

For energy these measures line-up closely with the those used by the current calibration. For mass, the current definitions fail to capture the impact of the high-response tail, meaning that any individual jet would have a mass scale around 3-4% higher than the actual JMS (jet mass scale) would suggest. The spread is less commonly seen and roughly corresponds to the $2\sigma$ value for a normally distributed dataset. It is useful to examine how well the calibration is doing at reducing instances of badly measured jets and can explain variance not captured by the scale or resolution.

The calibration scale and resolution are calculated for the uncalibrated jets, jets calibrated using the standard procedure and jets calibrated using each trained network. The response before calibration is calculated as in (5.6), the response after




(a) Energy Calibration

(b) Mass Calibration

Figure 5.5: The jet (a) energy and (b) mass response before and after the standard ATLAS large-R jet calibration is applied.

calibration is calculated as

$$\mathcal{R}_t^{calib} = \frac{t_{calib}}{t_{truth}} = \frac{t_{reco}}{\mathcal{F}_{calib} \times t_{truth}} = \frac{\mathcal{R}_t}{\mathcal{F}_{calib}} \tag{5.9}$$

where $\mathcal{F}_{calib}$ is the calibration function from (5.1) for the standard calibration and the neural network output $\mathcal{O}^{NN}(\vec{v})$ for the neural network calibration. In the rest of this text $\mathcal{R}_t^{uncalib}$, $\mathcal{R}_t^{calib}$ and $\mathcal{R}_t^{NN}$ will be use to refer to the response for uncalibrated, ATLAS-calibrated and neural network calibrated jets respectively.

Extensive plots were also made to understand the behaviour of the neural networks and their performance as a function of their inputs. Neural-networks have a reputation for being black-boxes as it can be difficult to understand the relationship between its inputs and output but we attempt to mitigate this through extensive profiling of each network. The behaviour of each network was profiled with respect to each input variable to find how both jet response and neural network output vary as a function of the network's input variables. In a ''perfect'' calibration these distributions would be identical, but since much of the jet's response is due to stochastic factors, the network's output is generally confined to a space close to the mean of the response in each bin. By taking the ratio of the input response to the network output, we obtain a distribution indicating the calibrated variables remaining dependence of the input distribution. The final measure of performance we use is the closure of the jet response distribution after calibration for each input variable.

### 5.2.2 Datasets

Network training and initial performance measurements were based on a sample of hadronic jets generated using Pythia8 with the A14 tune [113] and NNPDF23LO PDF set [114] then run through a GEANT4 simulation of the ATLAS detector [115] and the full ATLAS reconstruction software pipeline. The data was generated in slices of truth jet $p_T$, labelled ''JZ$x$W with $x \in [2, 12]$ with leading truth jet $p_T$ in the range $200 < p_T < 4600$ GeV. In each event jets were constructed using the standard ATLAS large-R jet algorithm, anti-$k_T$ with $R = 1.0$, and trimmed using $R = 0.2, f_{frac} = 0.05$. A copy of each jet was made with calibrated energy, $p_T$, and mass using the current ATLAS calibration, including the combined mass, calculated as described earlier in this chapter. The training was performed on the uncalibrated jets and these calibrated copies were solely for comparisons to the current calibration. For clarity, any references to a variable that do not explicitly refer to it as calibrated are referring to the uncalibrated variable. Truth and track jets are constructed using the same algorithm, taking Monte-Carlo truth particles and inner detector tracks, with $p_T > 400$ MeV and matched to the primary vertex, respectively as inputs. No calibration is applied to either the track or truth jets.

All events used were required to pass the following selection:

- $> 1$ truth, track and trimmed jet,
- No small-R jets labelled ''LooseBad'' by the ATLAS jet cleaning algorithm [116]

Each jet within an event is considered separately during the training and validation procedures, all trimmed jets within selected events passing the following criteria are considered:

- $p_T > 200$ GeV;
- $m > 30$ GeV;
- $|\eta| < 2.0$
- At least 1 truth and 1 track large-R jet with a separation of $\Delta R < 0.3$ from the jet and $p_T > 50$ GeV;

The substructure variables calculated with the truth and track jets are associated with the matched calorimeter jet. If multiple truth or track jets pass the $\Delta R$ separation requirement, then the one with the highest $p_T$ is chosen as the associated truth or track jet for that reconstructed jet.

The final sample consisted of 4 million jets taken at random from the totality of jets passing this selection and no sample weighting is performed (so the $p_T$ distribution of jets is not a realistic smoothly falling curve). This was then split into non-overlapping ‘‘training’’ and ‘‘validation’’ samples consisting of 3 million and ‘1 million jets respectively. The plots and performance figures shown in this study were made entirely with the validation sample.

#### Alternative Generator and Topology Samples

A range of further MC samples were used when evaluating the performance of the network in order to measure any generator or topology dependence of the trained calibration network. The additional datasets used were:

**Herwig QCD**: QCD multijet sample produced using Herwig++ and EvtGen. The samples were produced in identical slices to the Pythia QCD samples with approximately 2 million events per sample.

**WZ**: $W' \to WZ$ samples produced using Pythia8 and EvtGen. Samples were produced with $m_{W'}$ in the range 400 - 5000 GeV.

**Top**: $Z' \to t\bar{t} \to Wb$ samples produced using Pythia8 interfaced with EvtGen. Samples were produced at mass points between 400-5000 GeV with around 100,000 events per sample.

**Higgs**: $G_{RS} \to hh \to bbbb$ samples produced using MadGraph for the matrix element calculations interfaced with Pythia8 and EvtGen. Samples were produced at mass points from 400-3000 GeV with approximately 100,000 events per sample.

All samples used the A14 tune and NNPDF23LO PDF set, with the exception of the Herwig QCD sample, which used the CTEQ6.1 PDF set [117]. GEANT4 was used to perform the detector simulation for all samples. Jets from the Herwig dataset were selected using the same procedure as the Pythia jet sample and all jets in each event are considered. For the WZ, Top, and Higgs samples, the same selection is applied but only the two leading jets are considered, this ensures a relatively pure sample of bosonic/top jets.

#### Data Validation Samples

Finally, a sample of jets was taken from $\sqrt{s} = 13$ TeV, $pp$ data collected by ATLAS in 2016 during runs 300600-303291. Events were only considered if collected during normal running with nominal detector performance, corresponding to the Good Runs List: data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml. This gives a total luminosity of $7.16 fb^{-1}$, which was estimated by the methodology given in [118]. Any incomplete events, or those with bad LAr or Tile quality, were removed [119].

The data validation sample used the same selection procedure as the MC samples but to ensure good agreement with simulation only the leading jet in each event was considered. In addition, we required events to pass the 420 GeV large-R jet trigger. This trigger was not present in the MC samples and some shaping of the $p_T$ distribution occurs due to the large-R jet trigger turn on so a cut of 500 GeV was placed on the trimmed jet $p_T$ to reject any jets off the large-R jet trigger plateau. A Pythia8 QCD sample with these additional leading jet and $p_T$ cut requirements was generated to perform like-for-like comparisons between MC and data. Unlike the training and validation samples, MC event weights were taken into account and each jet $p_T$ slice was weighted to give the correct jet $p_T$ distribution. The MC sample was then scaled to match the luminosity of the data.

### 5.2.3 Input Selection

The choice of inputs to the proposed calibration network has a strong impact on both performance and the practicalities of training. We identified $\mathcal{O}(100)$ variables, including kinematic, substructure and constituent variables, which could conceivably contain information about the detector response for either the jet mass or energy. Unfortunately, simply training a network on all of these would be highly impractical for a number of reasons:

(1) The training time for a network greatly increases with the number of variables due to the increase in the number of input weights and the increase in network size necessary to handle additional inputs.

(2) Using a variable as input implies that we are confident it is simulated well by the Monte-Carlo (or at least that we can quantify the related uncertainty). Verifying that this is the case for such a large number of variables would be impractical.

(3) Many of the variables are highly correlated and adding an additional variable may not give the network any further relevant information.

(4) Many of the variables may not contain significant information about the detector response.

The existence of problems (3) and (4) makes (1) and (2) even worse. Including highly-correlated or unsuitable variables degrades training time without any improvement of network performance. The ideal input would be a minimal set of information-rich, independent variables. Previous jet substructure studies using machine learning have not treated this problem in any systematic way, typically a small set of candidate variables is chosen and the training is repeated using different combinations of variables until a seemingly optimal solution is found [106]. Since this would be impractical for a possible input space as large as the one examined, this study suggests two systematic methods for selecting variables which could apply to more general jet substructure studies.

### Variable Ranking

Picking an optimal input set given these criteria is non-trivial. Avoiding (4) requires some method of ranking the available variables based on how much information they contain about the detector response. An obvious metric for this would be the mutual information[3] between the target distribution $t$ and each variable distribution $v$, defined as

$$I(t;v) = \int_t \int_v p(t',v') \log\left(\frac{p(t',v')}{p(t')p(v')}\right) dt'dv', \tag{5.10}$$

where $p(t)$ is the probabilty density function (PDF) of $t$ and $p(t,v)$ is the joint PDF of $t$ and $v$. This can be calculated for the simulated data by approximating the PDFs using histograms. The exact value of $I$ was found to be sensitive to both the number of jets with which it was calculated and the histogram binning used in estimating the pdf. It was found that using a 200x200 bin histogram and more than 1 million jets to calculate each joint PDF gave stable rankings of the variables. The mutual information between each variable and the mass and energy response was calculated and two sets of variables chosen for each; referred to as the MI5 and MI10 input sets. These are the top five and ten ranked variables respectively after

[3] A dimensionless measure of the *total* shared information between two observables. It is substantially more general than correlation which captures only linear relations. To our knowledge this has only previously been used in jet substructure in the context of quark/gluon discrimination [120].

removing any variable whose mutual information with an already chosen (i.e. higher ranked) variable is greater than 1.5.

However, while this provides ranking criteria for our potential variables, it does not address the problem of correlations between them. The mutual information between each variable could be calculated, but there is no guarantee that this contains any information relevant to the jet response. Another option would be the n-variable generalisation of mutual information, but this would have required several orders of magnitude more data to approximate the n-variable PDFs.

### Iterative Principal Component Selection

One way to remove (linear) correlations from the dataset is to use principal component analysis (PCA), which maps the existing variables to the set of orthogonal variables with maximum variance. After mapping the new variables can be written as a weighted linear sum of the originals:

$$\tilde{v}^j = \sum_{v \in V} w_v^i v \tag{5.11}$$

where $w_v^i$ are the weights of a linear mapping from the old to the new variable sets. Note that the number of principal components found is not necessarily the same as the number of variables in the original input space.

Unfortunately, including the entire input space in the PCA procedure exacerbates problem (2); we require knowledge of the entire input space to assess the uncertainties of each new variable.

To address these problems we developed a method, referred to here as *iterative principle component selection* (IPCS) to find the fixed size set of orthogonal variables which maximises the mutual information while using as few input variables as possible. To do so, we define an informal measure of the theoretical maximum performance of a set of $N$ variables $\{v^0, \ldots, v^N\}$ for calibrating target $t$ as

$$P(V) = \sum_{i=0}^{N} I(t; v^i) \tag{5.12}$$

The procedure to find the optimal $N$ orthogonal variables using at most $M$ variables from the original normalised variable set $V$ is as follows

1. Calculate $I(t; v)$ for each $v \in V$ and rank from highest to lowest $v^0, v^1, \ldots$.

2. Select $\{v^0 \dots v^{M-1}\}$, as the basis set $V_B$.

3. Perform PCA on $V_B$ to find the orthogonal variable set $\tilde{V}$ and once again rank according to $I(t;\tilde{v})$. We take the $N$ top variables and call this the baseline orthogonal variable set $\tilde{V}_{baseline} = \{\tilde{v}^0, \dots, \tilde{v}^{N-1}\}$.

4. Define the baseline performance $P_{baseline} = P(\tilde{V}_{baseline})$.

5. Iterate over $V_B$ and remove each variable in turn, then compute $\tilde{V}$ and $P(\tilde{V})$ for each resulting input set. Take the reduced input set with largest $P(\tilde{V})$ as the new candidate set.

6. If $P(\tilde{V}_{candidate}) < (1-\epsilon)P_{baseline}$, then stop and, if it has been defined, take $V_{opt}$ as the set of orthogonal input variables. Otherwise, use $\tilde{V}_{baseline}$.

7. Define the new optimal variable $V_{opt} = \{\tilde{v}^0, \dots, \tilde{v}^{N-1}\}, \tilde{v}^i \in \tilde{V}_{candidate}$ and return to step 5.

This will find the basis of orthogonal variables that best preserves the total contained information while minimising the number of original variables used. The error parameter $\epsilon$ controls the maximum allowed reduction in the mutual information, which is always checked relative to the initial performance $P_{baseline}$. Note that we do not calculate $P$ for the non-orthogonal dataset, this is because the linear correlations between the variables, which is removed by PCA, make the comparison between pre and post PCA values meaningless. Additionally, we allow the basis set to be larger than the input set. This makes it possible to deal with problems (1) and (2) separately, network training time can be controlled by fixing $N$, and Monte-Carlo uncertainties can be controlled by fixing $M$. Another positive feature of this procedure is that the information contained in many original input variables can be concentrated into a smaller number of final ones, reducing training time and the requirements on network size. For this study, $M$ and $N$ were chosen to be 30 and ten respectively; this choice was made primarily on the computing power available.

**Final Input Sets**

The two selection procedures proposed thus far, mutual information (MI) and IPCS, give us adequate scope for examining the potential of multivariate calibrations, but it is also interesting to look at how neural networks perform as modelling methods when given the same information used in the current calibration. In total, five input variable sets were picked for testing:

| Input Set | Energy Inputs | Mass Inputs |
|---|---|---|
| Simple | $p_T$, $\|\eta\|$, $E$ | $p_T$, $\|\eta\|$, $m$ |
| TA | - | $p_T$, $\|\eta\|$, $m$, $m^{TA}$ |
| MI5 | $p_T$, $\|\eta\|$, $E$, $\tau_{3,wta}$, $C_2$ | $W^{track}$, $\tau_{1,wta}^{track}$, $\tau_{2,wta}^{track}$, $\tau_{3,wta}^{track}$, $C_2^{track}$ |
| MI10 | $p_T$, $\|\eta\|$, $e$, $\tau_{3,wta}$, $C_2$, $\tau_{2,wta}$, $p_T/m$, $N_{const}$, $W$, $\tau_{1,wta}$ | $W^{track}$, $\tau_{1,wta}^{track}$, $\tau_{2,wta}^{track}$, $\tau_{3,wta}^{track}$, $C_2^{track}$, $N_{track}$, $D_2$, $m^{TA}/m$, $W$, $\tau_{1,wta}$ |
| PA | $\|\eta\|$, $p_T$, $e$, $C_2$, $p_T/m$, $N_{const}$, $C_2^{track}$, $m$, $m_{TA}$ | $W^{track}$, $\tau_{1,wta}^{track}$, $W$, $n_{track}$, $C_2^{track}$, $W$, $p_T/m$, $m$, $A$, $C_2$, $ECF_2$ |

Table 5.1: Summary of the input sets used for training the energy and mass calibration networks.

SIMPLE - The existing calibration inputs, $p_T$, $|\eta|$ and either $E$ or $m$;

TA - The existing calibration inputs plus $m^{TA}$ (mass response only);

MI5 - The top 5 variables ranked by mutual information;

MI10 - The top 10 variables ranked by mutual information;

IPCS - Variables picked by iterative principal component selection with $M = 30$, $N = 10$ and $\epsilon = 0.02$.

The full variable sets are given in table 5.1.

The IPCS process was validated by measuring the performance $P(V)$ for each iteration, shown in figure 5.6. As expected, the performance of the ten selected principle components increased as less useful variables were excluded from the training set and the relevant information became concentrated into a smaller number of principal components. For the energy IPCS, the $P$ increased as the first ten variables were removed then plateaued as the next ten variables were removed. Once the number of remaining variables fell below the number of principal components, the performance began to fall rapidly. A similar picture was seen with the mass IPCS, though with a significantly shorter plateau.

The final principal component variables can be understood by examining the weights relating each component to the basis variables as shown in figure 5.7. Since all the basis variables were normalised to $\mathcal{O}(1)$, the weights for each variable should approximately reflect the importance of each basis variable to each component. A few principal component variables correspond almost exactly to basis variables,

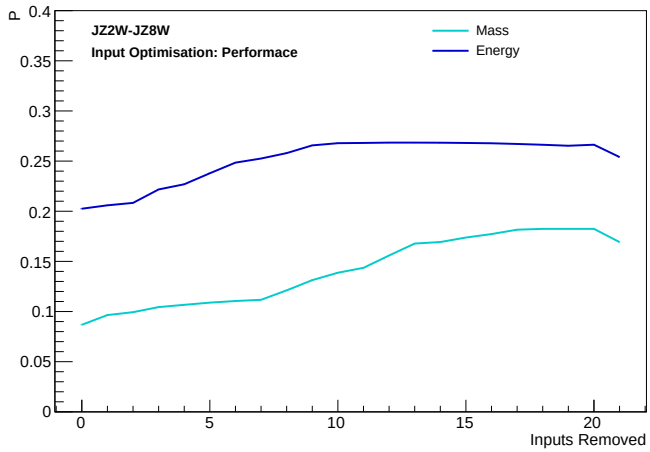



Figure 5.6: The performance, $P$, of the top 10 principle components during each step of the IPCS procedure for both mass and energy.

$p_T/m$ and $m$ for the mass response and $|\eta|$ for the energy response, but most are combinations of multiple basis variables.

### 5.2.4 Training

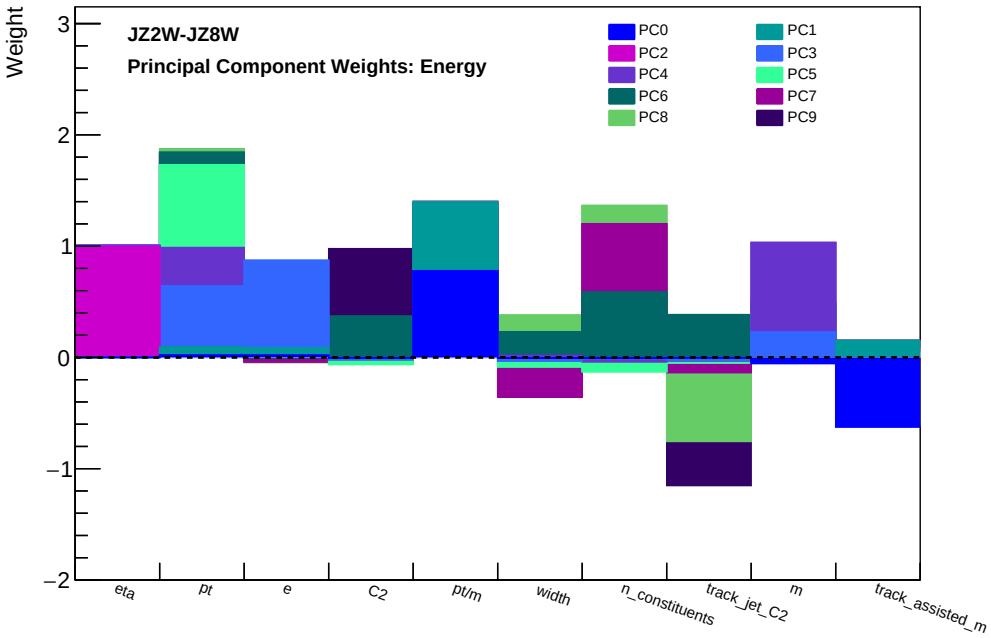
The calibration networks were implemented in python using Google's Tensorflow library [121]. Two loss functions were used during training; for energy the weighted-mean-squared-error (MSE) was used:

$$L(\vec{v}, \mathcal{R}_E, w) = \frac{[w \cdot (O_t(\vec{v}) - \mathcal{R}_E)]^2}{\sum w^2} \tag{5.13}$$

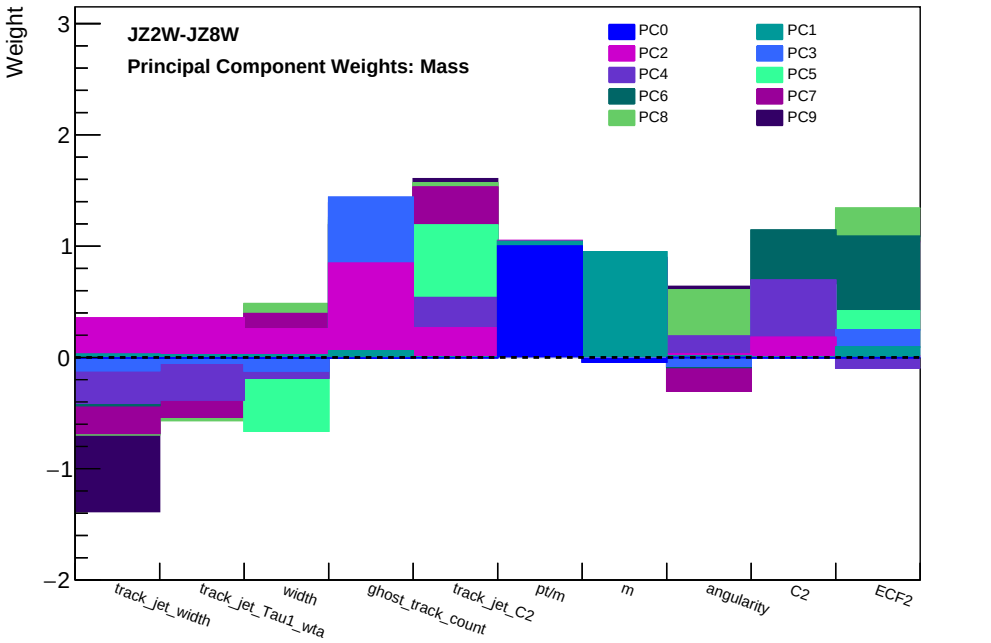
and for mass the weighted absolute error was used:

$$L(\vec{v}, \mathcal{R}_m, w) = \frac{|w \cdot (O_t(\vec{v}) - \mathcal{R}_m)|}{\sum w} \tag{5.14}$$

each for a jet with substructure variables $\vec{v}$, network output $O_t(\vec{v})$, true jet response $\mathcal{R}_t$ and a jet weight, $w$, which was set to 1. Different loss functions were chosen because MSE does not result in a calibrated response with $\hat{\mathcal{R}}_t = 1.0$ when the target distribution is strongly asymmetrical, as is the case for the mass response. This is due to the out-sized impact of the jets in the tail of the response which have very large squared-errors. Hence the mass calibration was trained with a linear loss function to reduce the impact of such jets.




(a)




(b)

Figure 5.7: The weights used in eqn. (5.11) to relate each basis variable to each principle component for (a) the energy response and (b) the mass response. Each plot shows the total weight assigned to each basis variable colour-coded by principle component.

To prevent over-fitting, dropout regularisation [122] with a frequency of 50% was used. This gives each neuron within a network a 50% chance of being turned off during each training step, preventing the network from developing a strong dependence on a small group of highly coupled neurons.[4] Multiple weighting schemes were investigated but using no weighting at all was found to give the smallest validation dataset error.

A ''flat'' weighting scheme, in which the jets were weighted to provide a flat $p_T$ spectrum, was tried but this was found to degrade the results significantly. The cause was believed to be large weights assigned to jets in certain $p_T$ regions whose substructure distributions were not reflective of the typical jet population. Other experiments with different weighting schemes showed little difference in overall performance, and as a result, no weights were applied during training or profiling of the networks. No sample weights (which would be needed to give a realistic $p_T$ spectrum) were applied, as these would overly bias the network towards the low $p_T$ region.

### Network Architecture and Hyperparameter Optimisation

In addition to the optimisation algorithm and error function, there are many tunable parameters involved in training a network: the network architecture, length of training, learning rate, and optimisation algorithm parameters. Though it is possible to optimise these parameters automatically, it requires substantially more computing power than was available [123] and as such the training hyperparameters were chosen through trial and error for each network.

The networks used for this study were all of the same basic design, consisting of an input layers, 1-5 hidden layers of ReLU[124] [5] neurons and a single output neuron which simply added its inputs. The networks were fully-connected, in that each layer receives an input from every neuron in the previous layer and sends an output to every neuron in the next layer. The network architecture was ''optimised'' by testing individual networks with different numbers of layers and neurons, two sets of 8 networks were tested, labelled the ''small'' and ''large'' networks sets. Each set consisted of a limited number of networks of varying depth and width, and the most complex network in the small set was around the same size as the smallest network

[4] In fact, rather than just avoiding overfitting, dropout has also been observed to increase network performance. An intuitive explanation for this is that the network approximates taking the average of an ensemble of smaller networks.

[5] A RectiLinear Unit, which use the activation function $f(x) = max(0, x)$

| Label | Architecture |
|---|---|
| S-1-50 | 50 |
| S-1-100 | 100 |
| S-1-200 | 200 |
| S-2-50 | 50-50 |
| S-2-100 | 100-100 |
| S-3-50 | 50-50-50 |
| S-3-100 | 100-100-100 |
| S-3-300 | 300-200-100 |

(a) Small Network Set

| Label | Architecture |
|---|---|
| L-1-50 | 500 |
| L-2-100 | 100-100 |
| L-3-100 | 100-100-100 |
| L-3-300 | 300-200-100 |
| L-4-100 | 100-100-100-100 |
| L-4-300 | 300-300-200-100 |
| L-5-100 | 100-100-100-100-50 |
| L-5-300 | 300-300-200-100-50 |

(b) Large Network Set

Table 5.2: The sets of network architectures used for testing. Each architecture is represented as the number of ReLU nodes in each hidden layer, with the input-connected layer on the furthest left and the output-connected layer on the furthest right. Each network set is subdivided by the number of layers in the network and each network is labelled by the total number of layers and the largest layer in the network.

in the large set. The full list of network architectures is given in table 5.2. The small network set was used for the Simple and TA input sets. The large network set was used for the MI5, MI10 and IPCS input sets.

Each network was trained on 3 million jets for a minimum of 1000 epochs, where an epoch is one cycle of training on each jet, in batches of 100,000 jets. The network was trained using a back-propagation style algorithm where the error in the network output for each jet was fed back into the network and the neuron weights updated based on this error, the speed of the training was controlled by the learning rate, which was set to 0.01. The back-propagation was performed using the Adam optimisation algorithm [125], which takes into account the momentum of the change in neuron weights alongside the current errors. The only Adam optimiser parameter tuned was the $\beta_2$ parameter, which controls the sensitivity of the training to the momentum of neuron weight changes, this was found to perform well for most networks with a value of 0.9.

## 5.3 Network Performance

After training, the performance metrics $\hat{\mathcal{R}}$, $IQR(\mathcal{R})$, and $IDR(\mathcal{R})$ were calculated on a validation dataset of 1 million jets, selected via the same procedure as the

| Input | Network | Architecture | Valid. Error | $\hat{\mathcal{R}}$ | $IQR(\mathcal{R})$ | $IDR(\mathcal{R})$ |
|---|---|---|---|---|---|---|
| Uncalib | - | - | - | 0.969 | 0.025 | 0.051 |
| Calib | - | - | - | 1.002 | 0.024 | 0.049 |
| **Simple** | **S-3-300** | **300-200-100** | **851.8** | **1.002** | **0.023** | **0.048** |
| TA | - | - | - | - | - | |
| MI5 | L-2-100 | 100-100 | 874.0 | 1.003 | 0.025 | 0.053 |
| MI10 | L-2-100 | 100-100 | 886.4 | 1.001 | 0.024 | 0.049 |
| IPCS | L-2-100 | 100-100 | 855.1 | 1.003 | 0.023 | 0.048 |

(a) Energy

| Input | Network | Architecture | Valid. Error | $\hat{\mathcal{R}}$ | $IQR(\mathcal{R})$ | $IDR(\mathcal{R})$ |
|---|---|---|---|---|---|---|
| Uncalib | - | - | - | 1.061 | 0.229 | 0.593 |
| Calib | - | - | - | 1.039 | 0.129 | 0.383 |
| Simple | S-3-100 | 100-100-100 | 3.32e6 | 0.996 | 0.126 | 0.386 |
| TA | S-2-100 | 100-100 | 3.14e6 | 1.002 | 0.117 | 0.361 |
| MI5 | L-3-300 | 300-200-100 | 3.12e6 | 0.989 | 0.116 | 0.338 |
| MI10 | L-3-300 | 300-200-100 | 2.77e6 | 0.995 | 0.098 | 0.284 |
| **IPCS** | **L-3-300** | **300-200-100** | **2.66e6** | **0.997** | **0.095** | **0.274** |

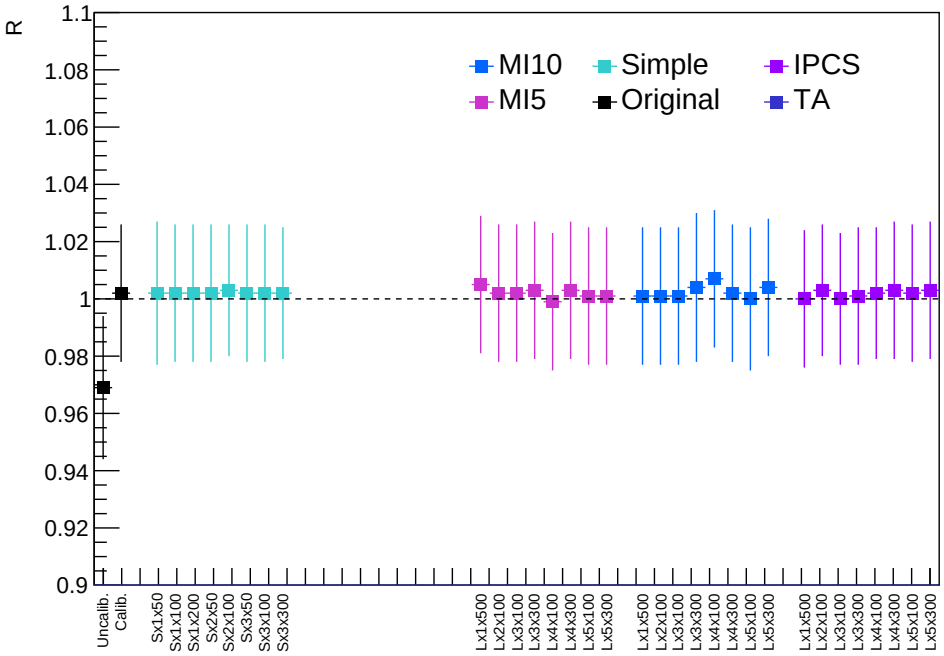
(b) Mass

Table 5.3: The calibrated jet scale and resolution for the best network in each input set for both (a) energy and (b) mass. The results for uncalibrated and standard calibrated jets are given for comparison. The ‘‘best’’ overall network for each of energy and mass is highlighted in bold.

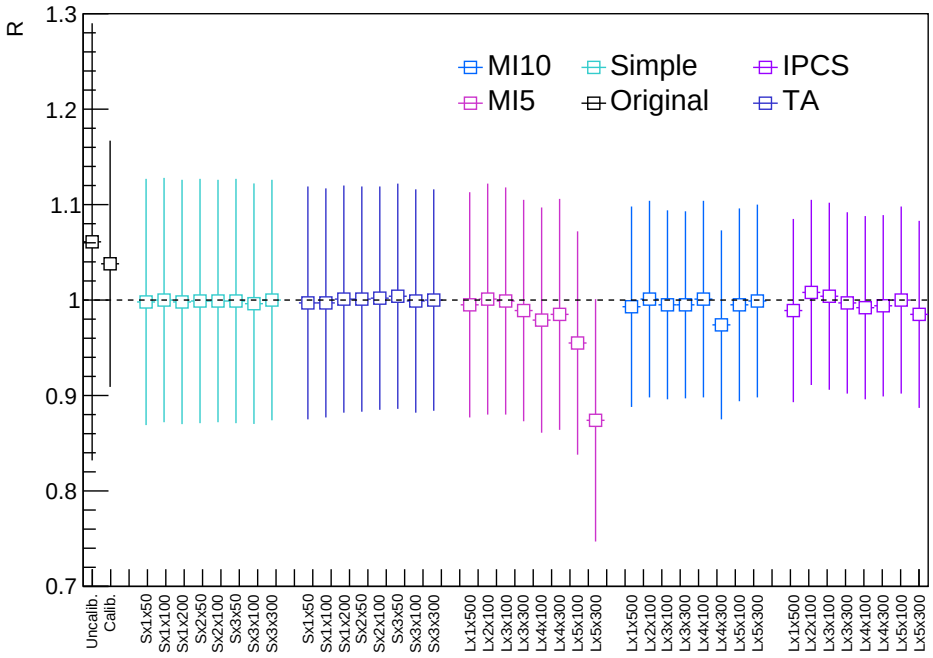
training dataset. The ‘‘best’’ network was chosen as the one with smallest total error on this dataset, using the respective MSE and absolute error functions for energy and mass. Figure 5.21 shows an overview of the results for each network and tables 5.3a and 5.3b show more detailed results for the best network in each input set for the energy and mass calibrations respectively. In the interest of brevity, the rest of this analysis will be confined solely to the best performing network for each of energy and mass. More extensive results are presented in appendix A.

### Energy Calibration Performance

For energy, it was found that performance was nearly identical both between each network and between the calibration networks and the standard calibration. The




(a)




(b)

Figure 5.8: Summary plots for the full range of energy (a) and mass (b) networks coloured by the network input. The median response and IQR of each network's calibration is shown by each point and its error bar. The uncalibrated and standard calibration results are shown on the left of each plot in black. Note that the energy calibration performance appears highly similar between each input type but the mass calibration shows a steady improvement in IQR as the number of inputs increases.

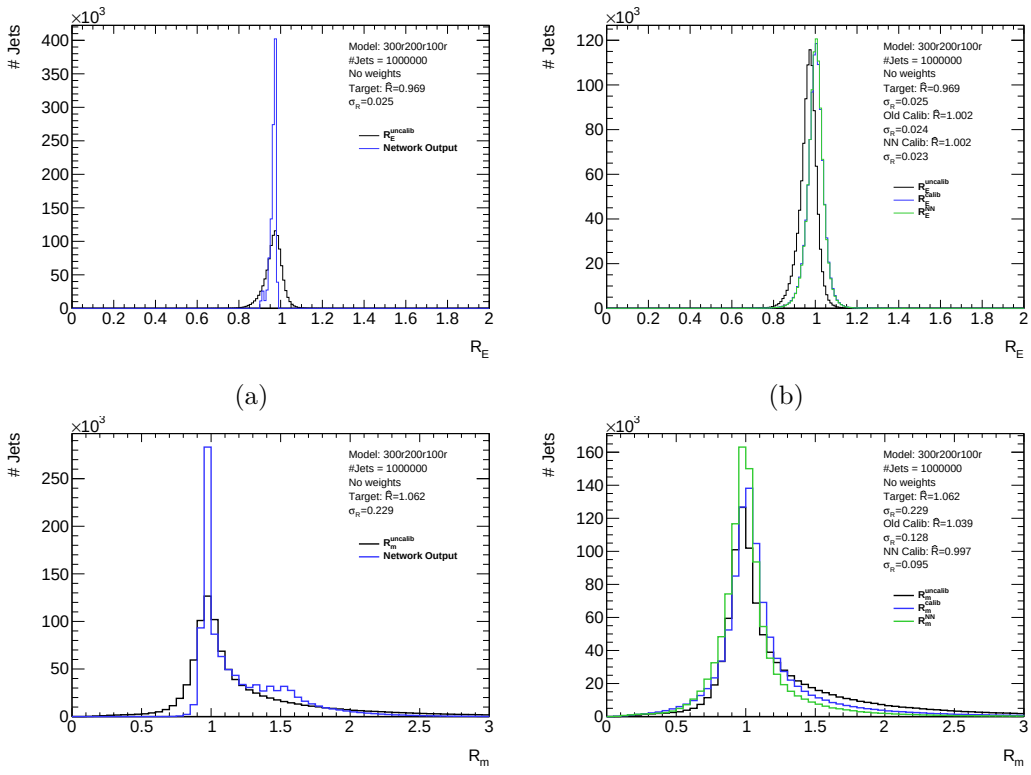



Figure 5.9: The energy and mass calibration network's performance on the validation dataset. Figures (a) and (c) show the distribution of the jet energy and mass response before calibration and the output distribution of the trained network on the validation dataset. (b) and (d) show the uncalibrated response alongside the original calibration and the neural-network calibrated response.

network with lowest validation error was network S-3-300 using the Simple input set. It can be seen in figure 5.9 that the response distribution is very close to the standard calibration but has a slightly taller peak. The calibration was found to have closure of $< 1\%$ in all bins of $p_T$, $E$ and $|\eta|$; with a resolution approximately equal to that of the standard calibration in each bin as shown in figure 5.10.

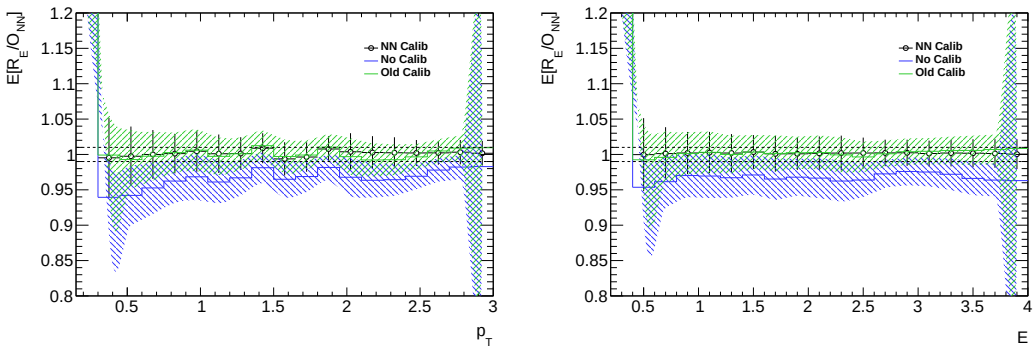
The fact that no significant improvement was seen for the Simple input set as network size increased suggests that the smallest network captured most of the information contained within the input set. This can be seen in figure 5.11 which shows an almost complete removal of any correlation with the mean jet response distribution as a function of each input variable. The lack of any improvement in performance from the addition of extra input variables indicates that the tested jet substructure variables may not contain any further significant information about the jet energy response.

One concern raised by figure 5.11 is the presence of structures in the mean response when plotted against $p_T$, and to a lesser extent $E$. These are caused by similar structures in the input $p_T$ distribution, which would ordinarily be corrected by the sample and event weighting of the Pythia QCD. The validation done in §5.3.1 and §5.3.2 shows that it does not have a significant impact on performance in other topologies but the data validation in 5.3.3 indicates more significant problems with these structures.

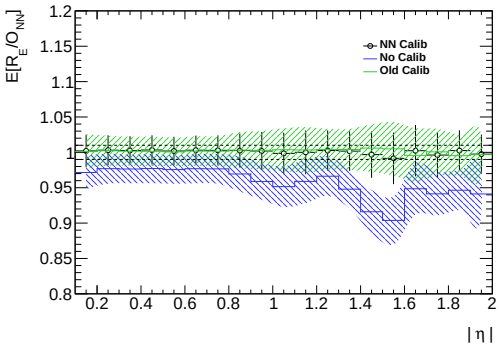
#### Mass Calibration Performance

The story is very different for the mass calibration network. Here we see multiple ways in which the network calibration is outperforming the standard calibration. Notably, the Simple network performs nearly identically to the calibrated combined mass despite having the disadvantage of not using $m_{TA}$ as an input. When $m_{TA}$ is added as an input in the TA dataset, we see a small improvement of around 8% in the resolution relative to the standard calibration. It seems reasonable to assume that this reflects a genuine improvement in the calibration procedure, at least for the proposed measures of quality.

The MI10 and IPCS datasets provide an even more substantial improvement. Both the resolution and spread of the best performing IPCS network were 26% smaller than the standard calibration, and the results were similar for all IPCS networks. The best MI10 network came very close to this figure, showing a 24% improvement in resolution. The mass network was profiled in the same fashion as the energy

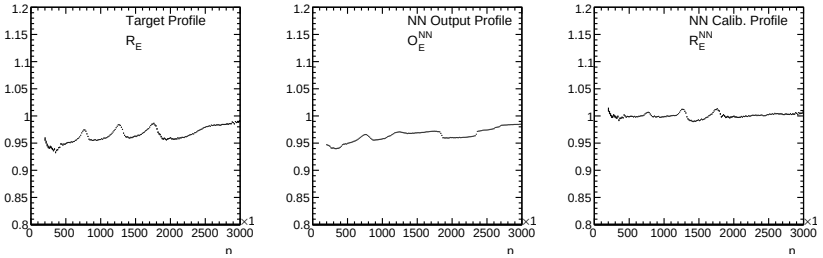



(a) $p_T$ (b) $E$

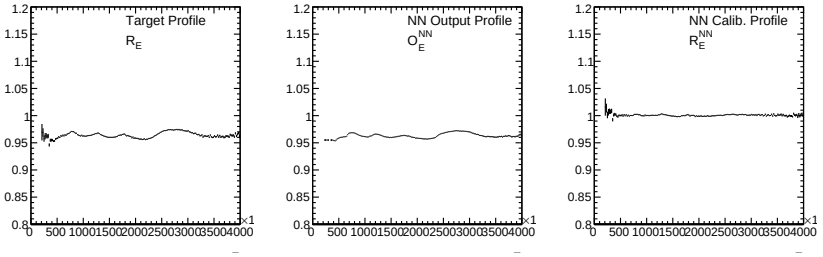



(c) $|\eta|$

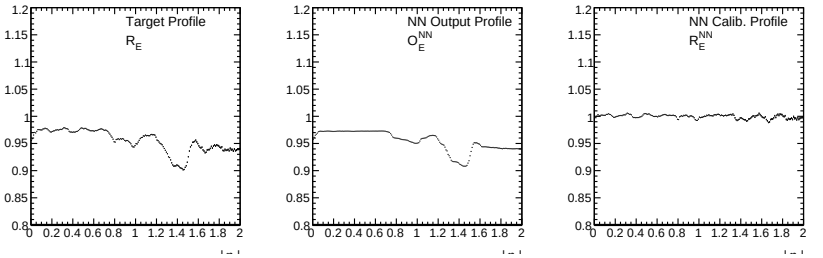
Figure 5.10: The closure of the energy calibration network and the original calibrations, in bins of each input variable. The blue and green solid lines show the uncalibrated and original calibration's $\hat{\mathcal{R}}_E$. The $IQR(\mathcal{R}_E)$ for each is indicated by the corresponding shaded area. The black line and errors bars correspond to the calibration network's $\mathcal{R}_E$ and $IQR(\mathcal{R}_E)$ respectively. The dashed line indicates a 1% distance from unity.




(a) $p_T$




(b) $E$




(c) $|\eta|$

Figure 5.11: Response profiles for each input variable to the energy calibration network. Each set of three figures shows, from right to left, the median of the jet response distribution, the median of the neural network output, and the median of the network calibrated jet response as a function of input variable, t. In a perfect calibration, the left-most and centre plot would be identical, with the right-most plot showing a flat line centred at 1.

network, focusing on the response to individual inputs, but despite using the IPCS input set, the network was profiled in terms of its raw inputs. Since there are a considerable number of inputs to this network, only the closure plots are presented in full. A complete overview of the results is presented in appendix A.2.

Figures 5.12-5.13 show the network closure as a function of its input variables, this allows us to observe the calibration performance in specific regions of the input space and identify any areas where the performance may be significantly improved (or degraded). From figure 5.13a it is easy to see that the improvement in mass resolution is across the whole range tested, but we can also find regions of phase space where the performance improvements are much greater. The resolution is notably improved at low values of $n_{track}$, $C_2$, $ECF_2$, and $W$ and the substantial variations at large values of $p_T/m$ have been removed. Of the inputs, only the angularity response seems to see no significant improvement.

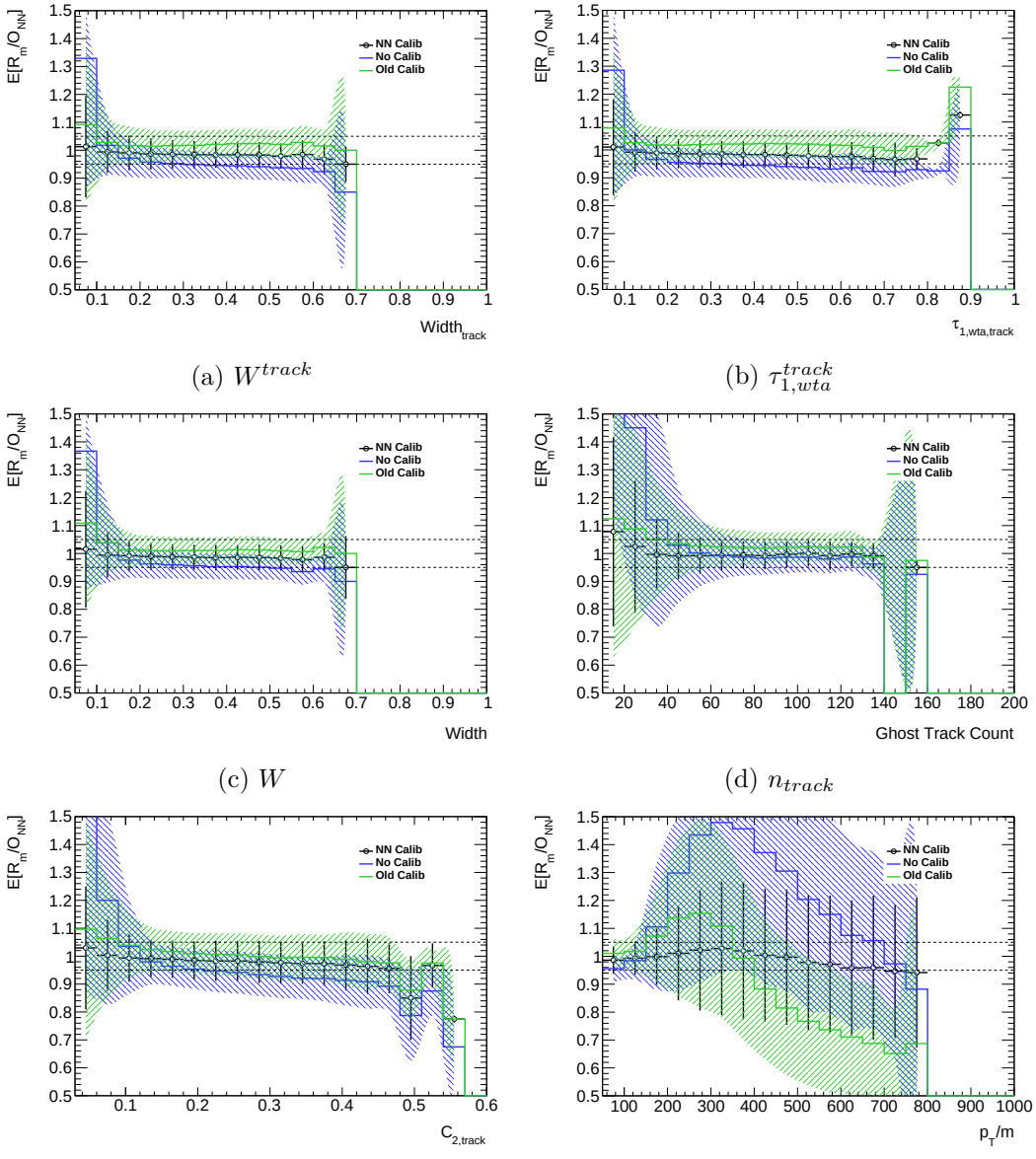
A subset of the response profiles are shown in figure 5.14. It is obvious from the post-calibration plots that the network is correctly learning the input distributions.

### 5.3.1 Generator Dependence

Since the networks were trained exclusively on a Pythia dataset, it is reasonable to assume that there may be some unwanted dependence on quirks of Pythia's shower and hadronisation models, rather than the actual detector physics. To attempt to quantify this effect, the network was tested on 1 million jets generated using the Herwig generator that pass the same cuts as applied to the original training dataset. Given the calibration should be most sensitive to the detector simulation rather than the MC generation, a good calibration should show only minor differences between generators. The performance of both energy and mass calibration networks on the Herwig dataset is shown in table 5.4.

For energy, the network was found to have $\hat{\mathcal{R}}_E^{NN} = 1.002$ and $IQR(\mathcal{R}_E^{NN}) = 0.023$ exactly equal to that obtained using the pythia validation set and closure for the Herwig dataset is shown in figure 5.15. Several small differences are observed but any discrepancies in closure after calibration correspond to discrepancies in the underlying uncalibrated dataset as would be expected.

The mass calibration network was also tested and while the difference in performance was small there was clearly some impact. The resolution and spread measured in the Herwig sample were $IQR(\mathcal{R}_m^{NN}) = 0.113$ and $IDR(\mathcal{R}_m^{NN}) = 0.341$, a 19% and

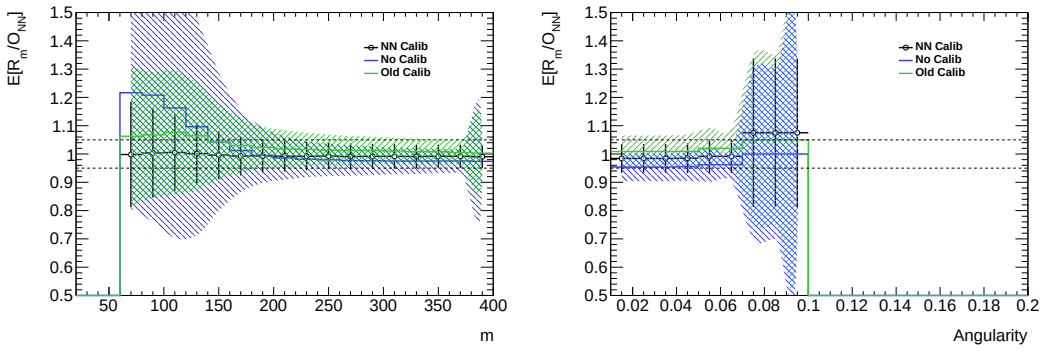



(a) $W^{track}$ (b) $\tau_{1,wta}^{track}$

(c) $W$ (d) $n_{track}$

(e) $C_2^{track}$ (f) $p_T/m$

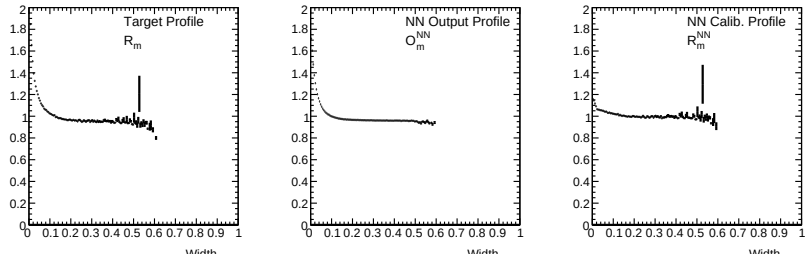
Figure 5.12: The closure of the mass calibration network and standard calibration in bins of each input variable. The dashed line indicates a 5% difference from unity. Continued in figure 5.13.



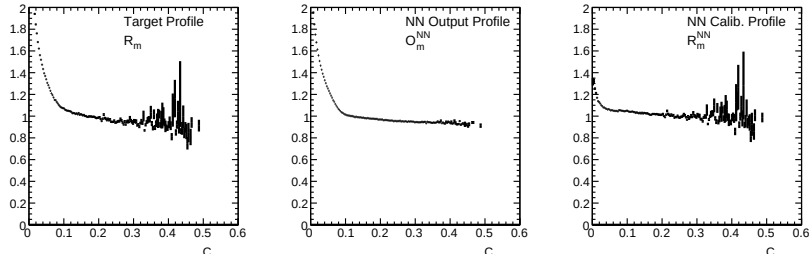

(a) $m$ (b) $A$

(c) $C_2$ (d) $ECF_2$

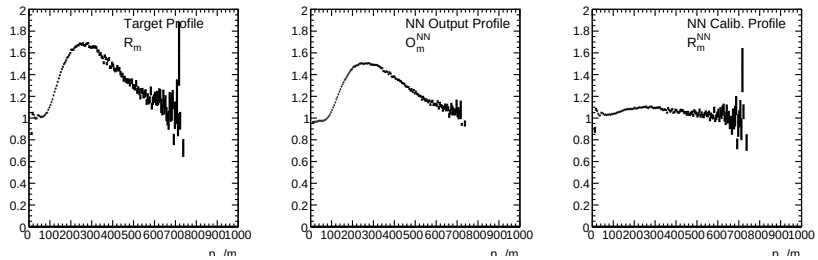
Figure 5.13: The closure of the mass calibration network and standard calibration in bins of each input variable. The dashed line indicates a 5% difference from unity. Continued from figure 5.12.



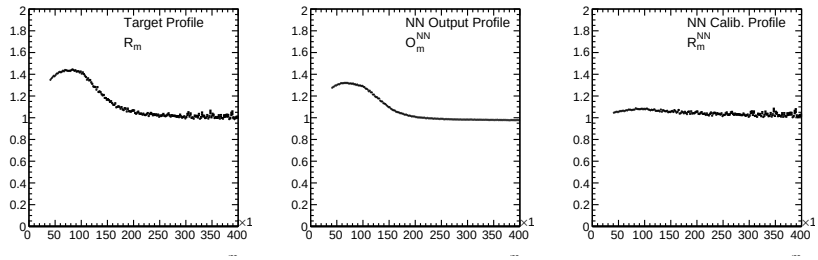

(a) $W^{track}$




(b) $C_2^{track}$

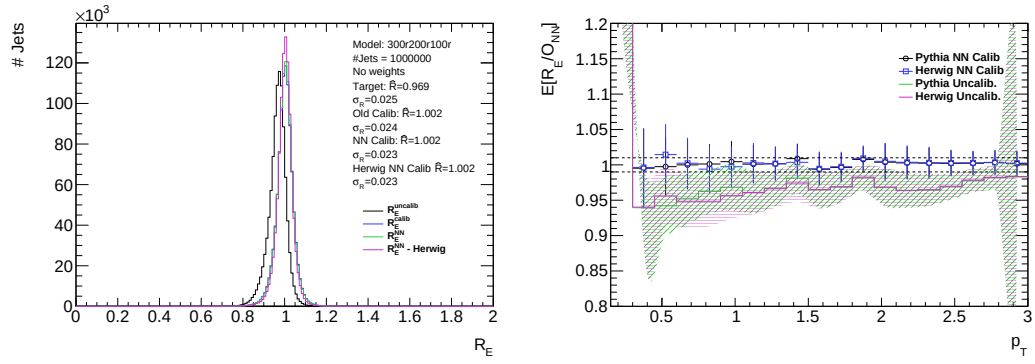



(c) $p_T/m$




(d) $m$

Figure 5.14: Response profiles for a subset of the input variables to the mass calibration network. For each variable the plots show (from left to right) the mean response, the mean calibration output, and the mean calibrated response as a function of that variable. The significant flattening of each response function post-calibration indicates much of the non-stochastic variation in the mass response has been removed. The large error bars seen at the ends of the $C_2^{track}$ and $p_T/m$ distribution is due to a lack of statistics in those regions. Plots for the full input set are shown in appendix A.



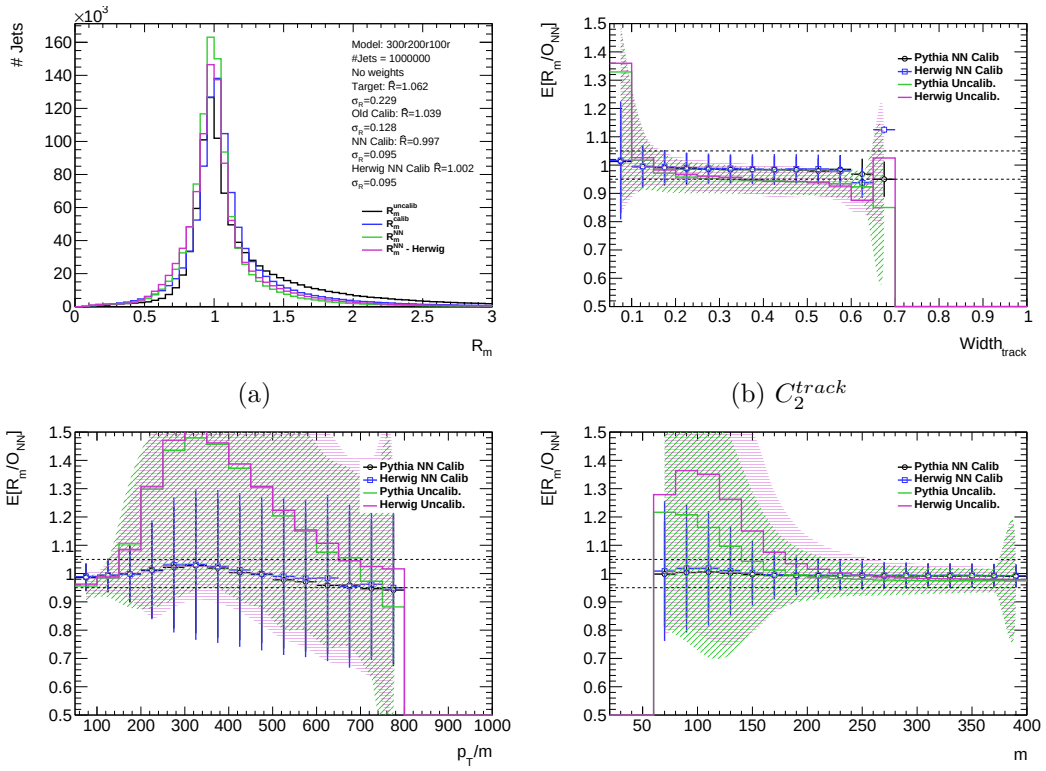

(a)

(b) $p_T$

(c) $E$

(d) $|\eta|$

Figure 5.15: A comparison of the energy calibration network performance when tested on jets from the Herwig QCD dataset and jets from the Pythia QCD validation set. Figure (a) shows the response for each dataset and the (b)-(d) show the closure in each input variable. Performance is found to be extremely similar in both datasets.




(a)

(b) $C_2^{track}$

(c) $p_T/m$

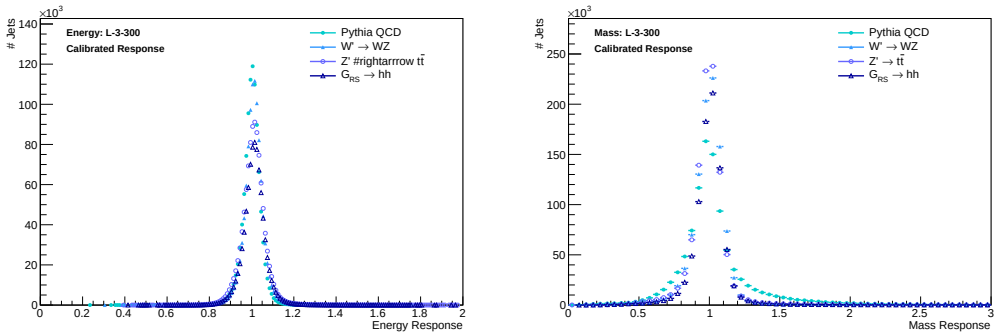
(d) $m$

Figure 5.16: A comparison of the mass calibration network performance when tested on jets from the Herwig QCD dataset and the Pythia QCD validation set. (a) shows the response for each dataset and the other plots show the closure in each input variable. Performance is found to be extremely similar between both datasets.

24% increase respectively. However, this must be taken in the context of the Herwig sample, in which the uncalibrated resolution and spread are 29% and 24% worse than that of the Pythia sample. The degradation in performance on Herwig samples is approximately equal to that experienced by the standard mass calibration. This suggests that any difference in performance is not due to the network learning quirks of the Pythia generator. The closure plots in figure 5.16 and the appendix show that the bin-by-bin differences between Herwig and Pythia are negligible.

### 5.3.2 Topology Dependence

Since we are considering jet substructure during this calibration procedure it is also useful to consider non-QCD jets from boosted vector-boson, top and Higgs boson decays. One of the theoretical advantages of a multi-variate calibration procedure is




(a) (b)

Figure 5.17: A comparison of the neural-network calibrated jet response for Pythia QCD, $WZ$, $t\bar{t}$ and $hh$ jets for the energy (a) and mass (b) calibration networks. A slight shift in the scale is observed for energy and the mass resolution is found to be slightly better for jets from heavy objets decays.

| | Energy Calibration | | | Mass Calibration | | |
|---|---|---|---|---|---|---|
| Dataset | $\hat{\mathcal{R}}$ | $IQR(\mathcal{R})$ | $IDR(\mathcal{R})$ | $\hat{\mathcal{R}}$ | $IQR(\mathcal{R})$ | $IDR(\mathcal{R})$ |
| Pythia | 1.002 | 0.023 | 0.048 | 0.997 | 0.095 | 0.274 |
| Herwig | 1.002 | 0.021 | 0.043 | 1.001 | 0.113 | 0.341 |
| $WZ$ | 1.010 | 0.025 | 0.052 | 1.002 | 0.062 | 0.129 |
| $t\bar{t}$ | 1.011 | 0.031 | 0.064 | 0.995 | 0.056 | 0.123 |
| $hh$ | 1.013 | 0.029 | 0.061 | 0.008 | 0.056 | 0.121 |

Table 5.4: The performance of the energy and mass calibration networks across all datasets.

that the differences between these types of jets should be captured by the substructure information fed into the neural network. To test this, the response and closure was calculated for jets from $WZ$, $t\bar{t}$, and $hh$ datasets that had passed the selection detailed in §5.2.2. The results for both networks for all toplogies are shown in table 5.4.

For the chosen energy calibration, much of this supposed advantage is lost due to not using additional inputs, and indeed this is reflected in the validation plots. While figure 5.17 does not show a significant difference in the response, figures 5.18, show that the jet energy scale is typically around 1% too high for these jets. This discrepancy is also observed in the standard jet calibration shown in figure 5.19 and hence is likely a difference in the underlying data.

The mass calibration fares substantially better and the mass scale remains within

1% of unity. However the mass resolution is found to be topology-dependent. The resolution for vector-boson, top and higgs jets is found to be $\sim 6\%$, 2/3rds that of the QCD jets. For the top and higgs jets this is consistent or only slightly better than the standard mass calibration, however, it is a 26% improvement in the resolution of vector-boson jets. Figure 5.20 shows the behaviour of the mass calibration when applied to these topologies.

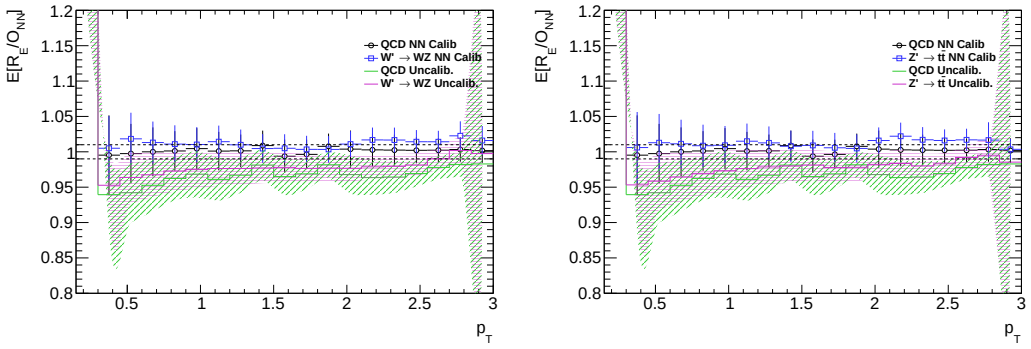
### 5.3.3 Data Validation

The calibration networks' reliance on simulated data for both training and validation raises the possibility that it may be learning unrelated elements of the simulation process rather than a correct calibration. Unfortunately, no ''truth'' variables are available in data so we cannot make straightforward measurements of the calibration's performance but it is still possible to validate the network performance in data by ensuring that its behaviour is comparable to simulation.[6]

Neural networks learn multivariate correlations between their inputs and target variable, which makes their behaviour difficult to understand. We attempt to asses both the energy and mass networks by examining the mean output of each as a function of their inputs, target variable, and $p_T$ and $\eta$ (when these are not already used as inputs). The logic behind this is as follows, if the mean network output, as a function of a single input variable, is consistent between two data samples then we can be confident that the correlations learnt by the network between all of the network's other inputs are, on average, consistent between the two samples in each bin of that initial variable. By checking this for each network input it can be proven that the correlations learnt by the network are also present in data, since non-matching correlations would give systematic disagreements between data and simulation. Note that this is not an assertion that the network output is correct, it is entirely possibly for the correlations between inputs to be correct but the correlation between those inputs and the network target (i.e. the actual correction factor back to a truth variable) to differ between data and simulation.

This is not a reversible argument; disagreement between the average network output in data and simulation is not, in-itself, proof of mismodelling in the network. Differences in the input data distributions can also lead to different outputs. To minimise this we have tried to ensure that the data and Pythia QCD samples match as closely

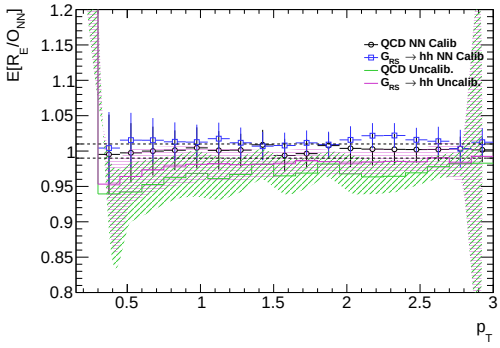
[6] That is not to say more precise measurements of the true jet energy or mass is not possible, $\gamma + j$ events, multijet balancing, and semi-leptonic $t\bar{t}$ have all be used to attempt to measure the performance of MC-derived calibrations [126] but these are beyond the scope of this study.



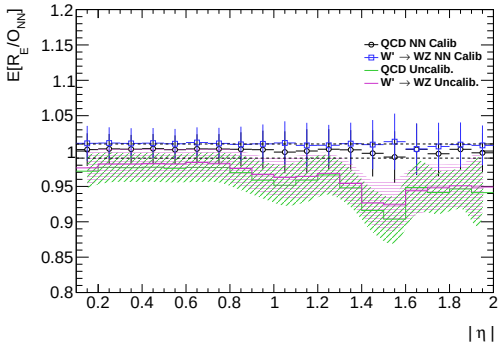

(a) $WZ$ - $p_T$



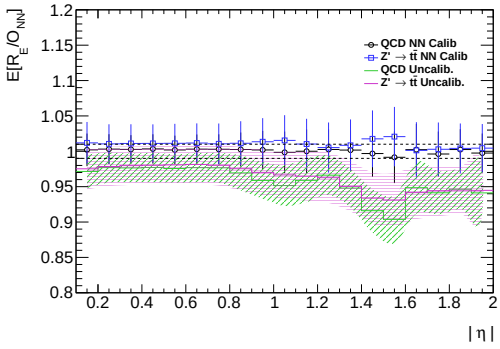

(b) $t\bar{t}$ - $p_T$



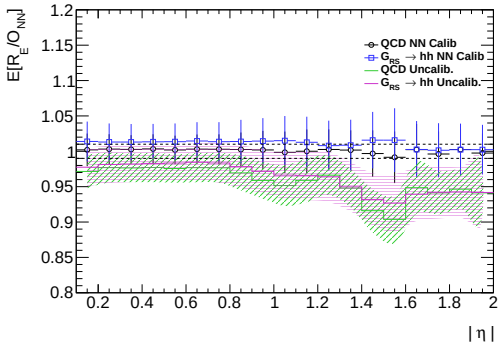

(c) $hh$ -$p_T$




(d) $WZ$ - $|\eta|$

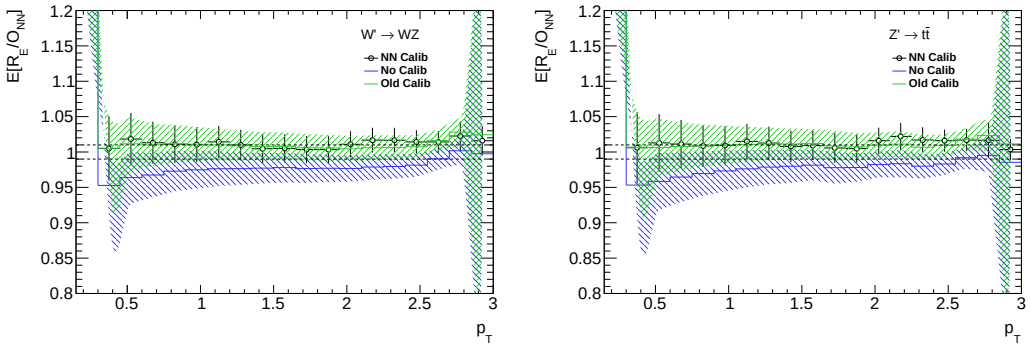



(e) $t\bar{t}$ - $|\eta|$




(f) $hh$ - $|\eta|$

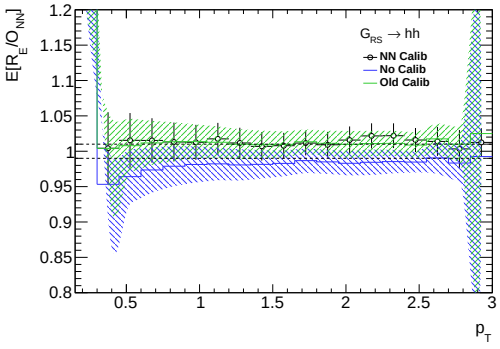
Figure 5.18: The closure of the energy response before and after the neural network calibration is applied for vector-boson, top and higgs jet samples in comparison to the Pythia QCD sample. The closure is shown as a function of $p_T$ in (a)-(c) and $|\eta|$ in (d)-(f).




(a) $WZ$

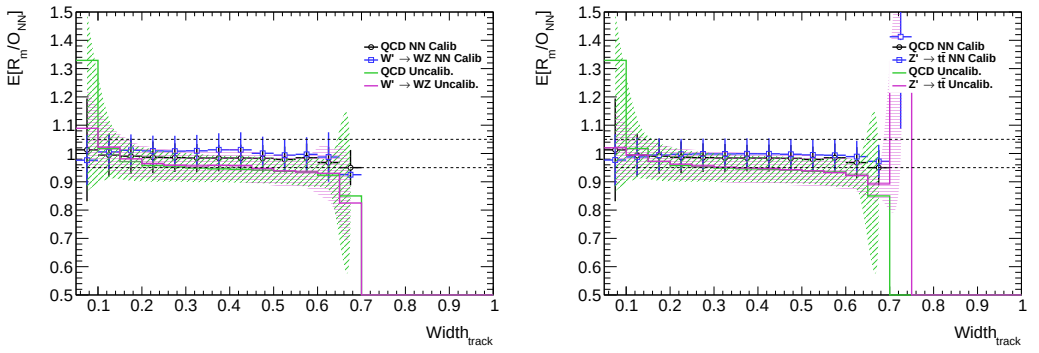



(b) $t\bar{t}$




(c) $hh$

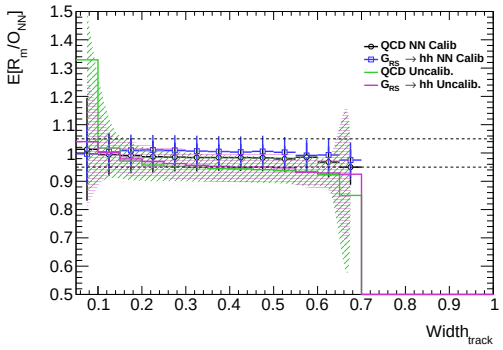
Figure 5.19: The closure of the neural-network and standard calibrations in bins of $p_T$ for $WZ$, $t\bar{t}$ and $hh$ samples. A systematic over-correction of around 1% is seen for each dataset in both the standard and neural network calibrations.



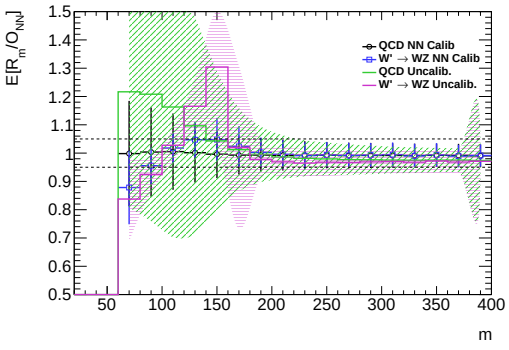

(a) $WZ$ - $W^{track}$



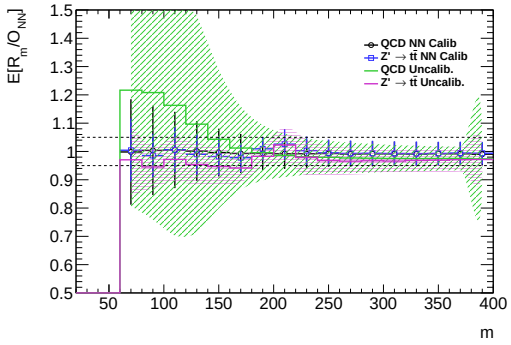

(b) $t\bar{t}$ - $W^{track}$



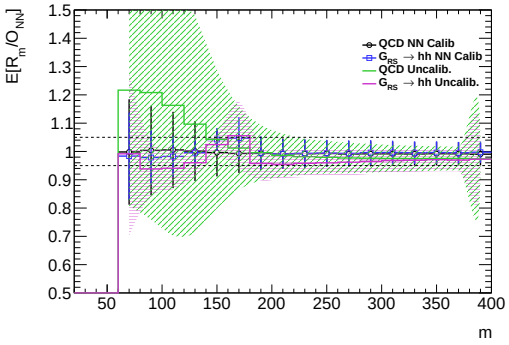

(c) $hh$ -$W^{track}$




(d) $WZ$ - $m$




(e) $t\bar{t}$ - $m$




(f) $hh$ - $m$

Figure 5.20: The closure of the mass response before and after the neural network calibration is applied for vector-boson, top and higgs jet samples in comparison to the Pythia QCD sample. The closure is shown as a function of $W^{track}$ in (a)-(c) and $m$ in (d)-(f). The closure as a function of the other network inputs is shown in appendix in figures A.7-A.12.

as possible by weighting the Pythia QCD sample to a data-like $p_T$ distribution. After reweighting there is excellent agreement in $p_T$ and mass between the data and simulated samples but there are substantial differences in the the $\eta$ and energy distributions. The increasing discrepancy at high energy is due to the cut $p_T < 3000$, which becomes angularly dependent in energy (note, this is the full energy, not $E_T$) as shown in figure 5.22. Hence the deficit of MC events at high $|\eta|$ is reflected as a similar deficit at high energy.

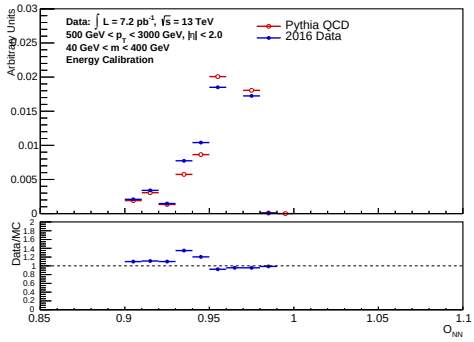
One additional observation about the reweighting process is that the reweighted Pythia sample is itself a validation sample. Any bias due to the frequency of jet behaviours within the training sample would be manifest in both the reweighted Pythia QCD and the data sample.

Figure 5.23 compares the data and MC distributions of all energy inputs, and the energy calibration network outputs as a function of each variable. The agreement between data and simulation as a function of all inputs is excellent which indicates the network correctly learnt physical correlations between its input variables. Unfortunately, examining the calibrated $p_T$ distribution raises some concerns. There are $p_T$-dependent structures in the network output, which appear in both simulation and data, that cause similar structures to appear in the calibrated $p_T$ distribution. These are also apparent in the network output as a function of the jet energy. The structures appear to occur at the boundaries of the QCD jet slices which formed the training and validation set, suggesting that the uneven nature of the training sample has caused problems with the network.

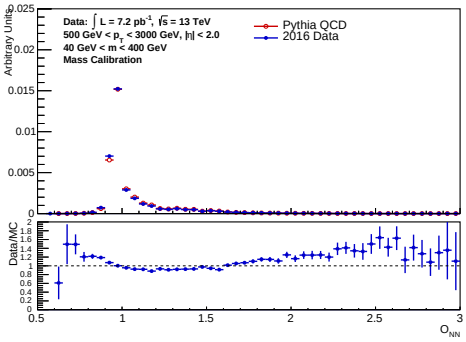
It is likely the case that the network has become biased towards jet behaviours which are overrepresented on the edges of the Pythia sample's $p_T$ slices but suppressed in data. When sample and MC event weights were applied to the Pythia validation sample this caused the appearance of non-physical structures in the network output. The matching impact on MC and data tells us that, despite the network having correctly learnt the correlations between inputs variables, the energy calibration network provides incorrect values for the energy correction factor.

This was not observed in the previous validation samples as both training and validation samples would see these type of jets with the same frequency. Any future work on calibrating energy using these networks would need to address how to remove these structures without suffering the loss in network capability seen when flattening the input $p_T$ distribution.

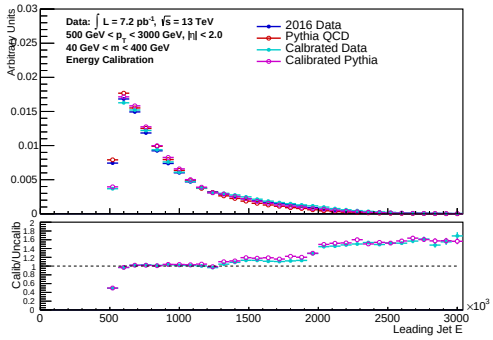
No such structures are seen in the mass calibration network, for which there are



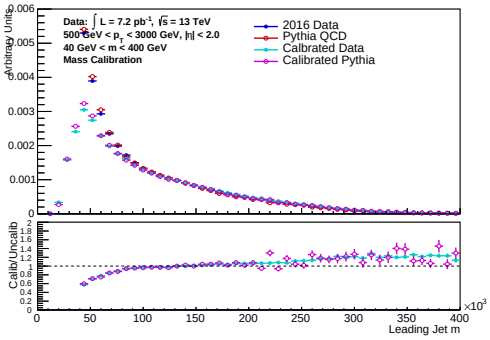

(a) $O_{NN}$ - Energy




(b) $O_{NN}$ - Mass




(c) Calibrated Energy




(d) Calibrated $m$

Figure 5.21: Network output and the pre and post-calibration mass distribution in data and MC. Network output agreement as a function of the target variable (the ratios shown in (c)-(d)) is excellent for both energy and mass networks but the differences in the data and MC energy distribution causes the shape difference in the total network output shown in plot (a). Structures have been created in the energy distribution (c) post-calibration for both simulation and data but none are observed in the mass calibration.

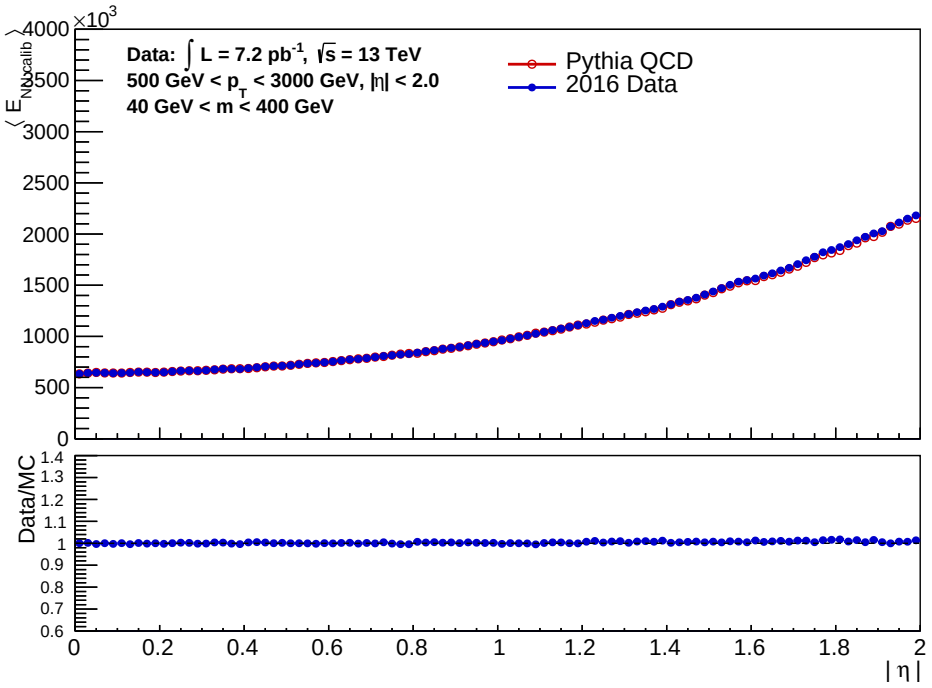
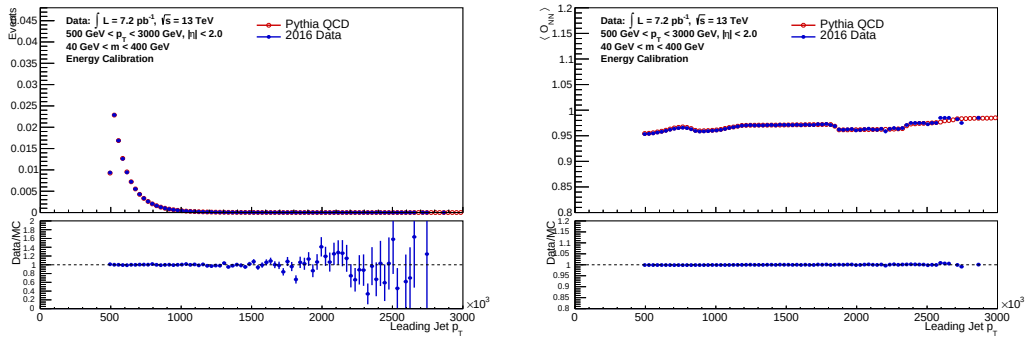


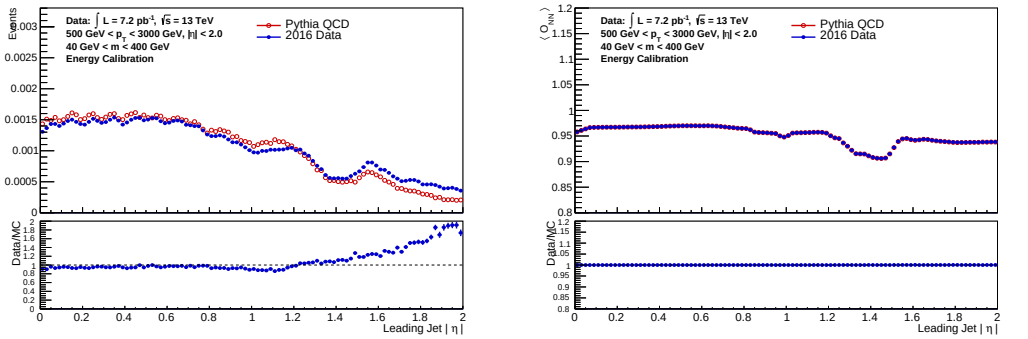

Figure 5.22: The $\eta$ dependence of jet energy in MC and data. Good agreement is seen between samples.

several possible explanations. Firstly, the mass network has the same number of neurons but a larger number of inputs, making overtraining more difficult. Additionally, jets on the edge of a $p_T$ slice (ie. reconstructed with a $p_T$ relatively far from their true $p_T$) are the result of actual physical behaviour which can likely be at least partially captured by their substructure. Since the mass network uses substructure information this type of jet behaviour could be captured by the large input parameter space, correctly accounting for this without impacting other types of jet. Finally, the mass distribution is less affected by the $p_T$ slicing than the energy distribution. All three of these factors are likely involved in the full answer but further study would be needed to know more. At the very least, the lack of structures associated with jet $p_T$, shown in figure 5.24, gives confidence that the mass calibration network does not exhibit evidence of overtraining on the weighting (or lack of) of the training sample.

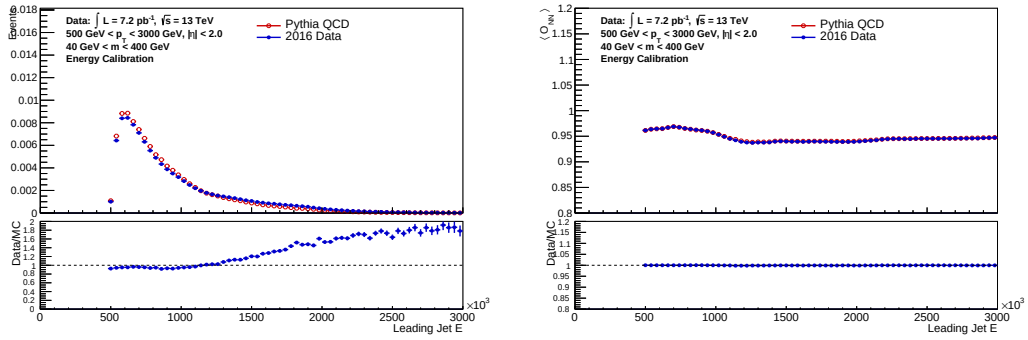
Figures 5.25 - 5.28 show the variable and network output distributions for the mass network's inputs. The agreement between variable distributions in data and MC is typically excellent, with the exception of $n_{track}$, which shows a small tendency to lower values in data than in simulation (this is studied more extensively in §7.6.4




(a) $p_T$

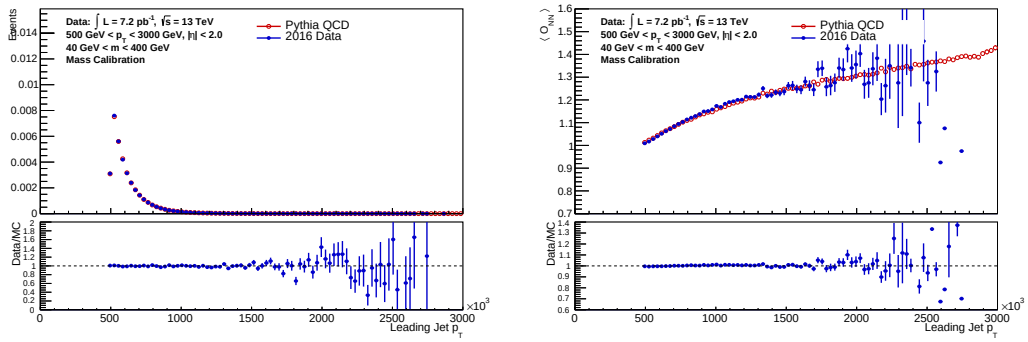



(b) $|\eta|$

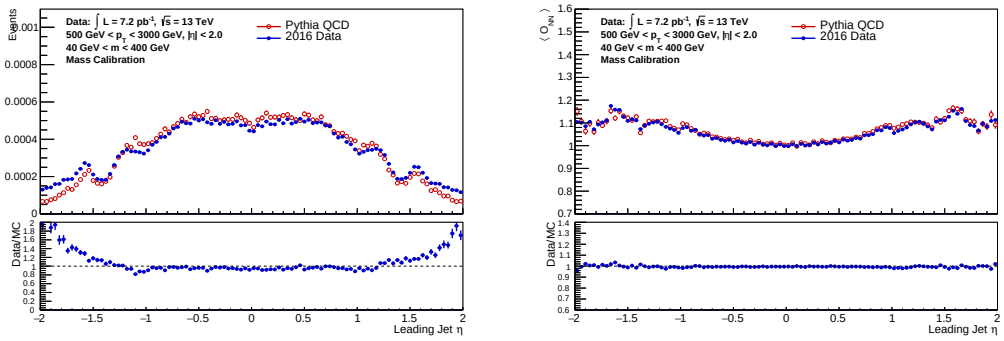



(c) *Energy*

Figure 5.23: The plots on the left show a comparison of each energy calibration network input variable in the data and MC samples. The right plots show the calibration network output as a function of these variables. Note that even when the variable distributions differ significantly between data and MC the network output is highly consistent, with differences in mean $O_{NN}$ smaller than 1% across the whole range of all variables.




(a) $p_T$




(b) $\eta$

Figure 5.24: The data/MC distributions and mass calibration network output as a function of the jet $p_T$ and $\eta$. Note that neither of these variables are inputs to the network so all effects of the calibration are due to their correlations with other network inputs. Good agreement is observed in the $p_T$ distribution but the MC is too central in $\eta$. Despite this, good agreement is achieved between the network output in MC and data as a function of both variables.

and found to be a difference of 1 track (±3%) between data and MC), and in low values of $C_2$, $C_2^{track}$, and $p_T/m$, where discrepancies of up to 20% are observed. Fortunately, these differences are away from the bulk of each distribution. Much smaller differences are seen in the network output. Differences between data and MC in the median network output as a function of each variable are typically smaller than 1%, including in regions where much larger discrepancies are seen between the input distributions. The largest differences seen are in low regions of $C_2^{track}$ where there is a difference of around 5% between the data and MC sample network output.

#### Pileup Depedendence

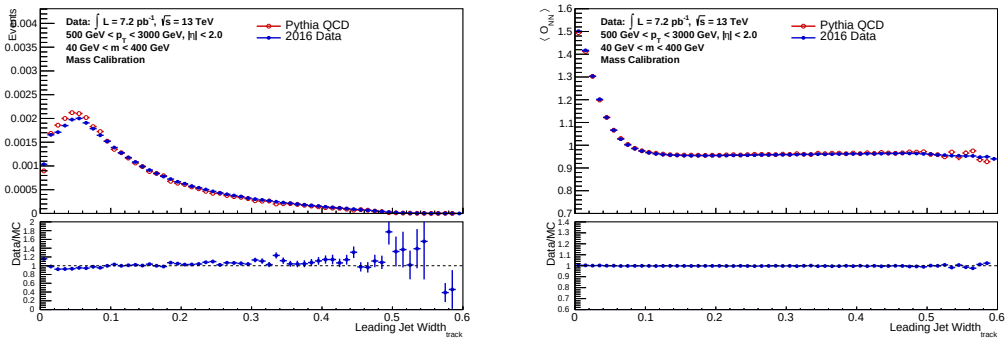
The jet trimming procedure aims to minimise the effect of pileup on jet variables and, as shown by figure 5.4(b), is reasonably successful at this. However, it is possible that some of the input variables used may still have some residual dependence on pileup. This is particularly important for this study as the pileup distribution in data was higher than the in the MC samples as shown in figure 5.29. The mean network output as a function of average pileup was plotted in figure 5.30 and both energy and mass calibration networks were found to have no significant pileup dependence.

## 5.4 Conclusions

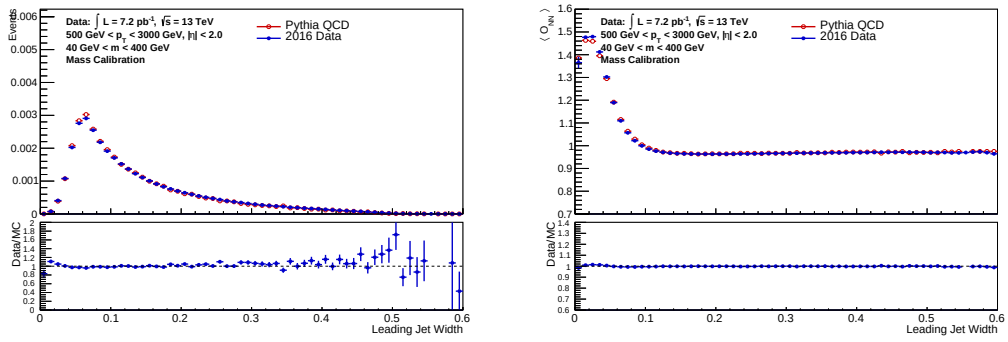
The primary aims of this study were to provide an indication of whether a neural-network-based large-R jet calibration is feasible, and to explore systematic methods for deriving calibrations using jet substructure. To produce the final two networks, a total of 72 networks were trained from 15 network architectures, and nine input sets.

Two methods of selecting input variables were proposed as ways to systematise variable selection from the extensive library of substructure variables: mutual information ranking and the iterative principal component selection algorithm (IPCS). For the mass calibration, both of these proved successful. Figure 5.31 gives an idea of how much the network performance varied within each dataset and it is interesting to note that the IPCS algorithm consistently out-performed the MI10 variables, with only one non-IPCS network coming close, despite being trained on the same network architectures and with the same number of inputs.

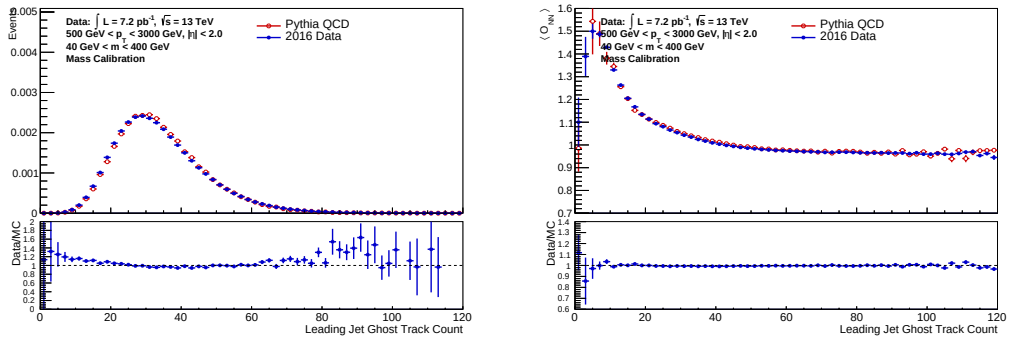
The performance of energy-calibrating-networks was at best equal to, and usually




(a) $Width^{track}$

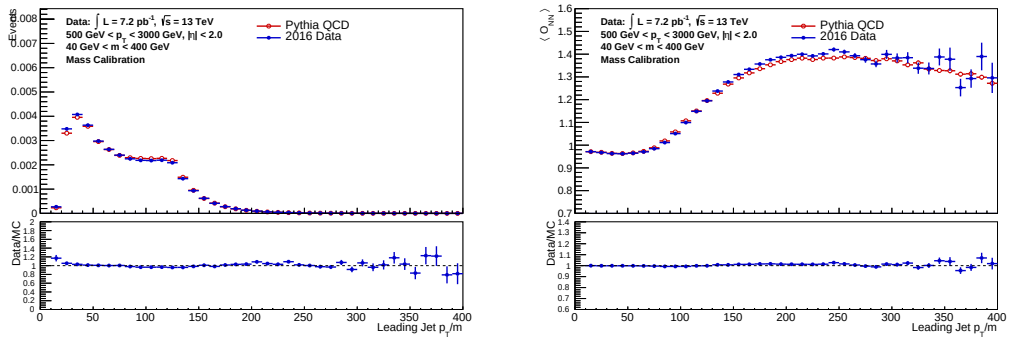



(b) $Width$

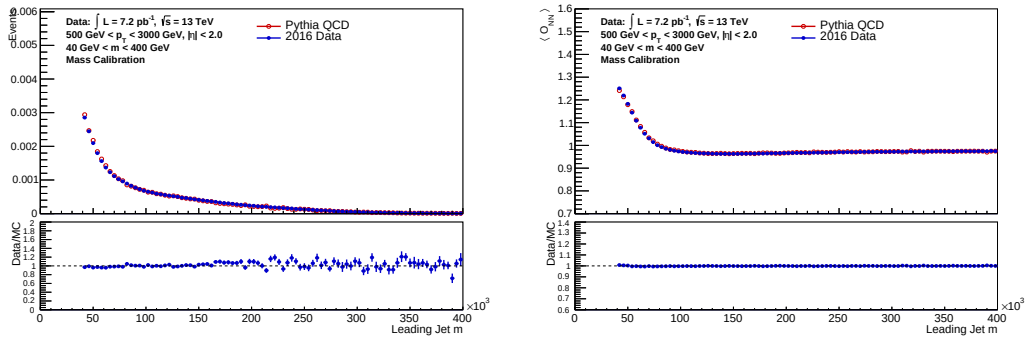



(c) $n_{track}$

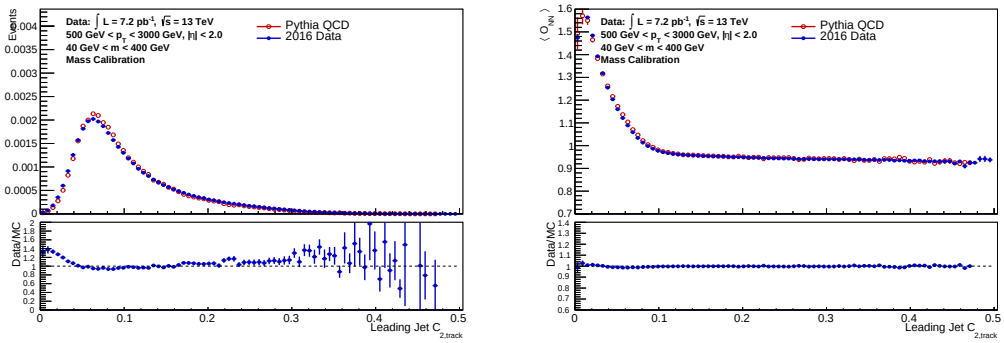
Figure 5.25: The plots on the left show a comparison of the $Width^{track}$, $Width$, and $n_{track}$ mass calibration network input variables in the data and MC samples. The right plots show the calibration network output as a function of these variables. The data/MC distributions all show good agreement with the sole exception of $n_{track}$. Network output agrees to $< 1\%$ for all variables.




(a) $p_T/m$

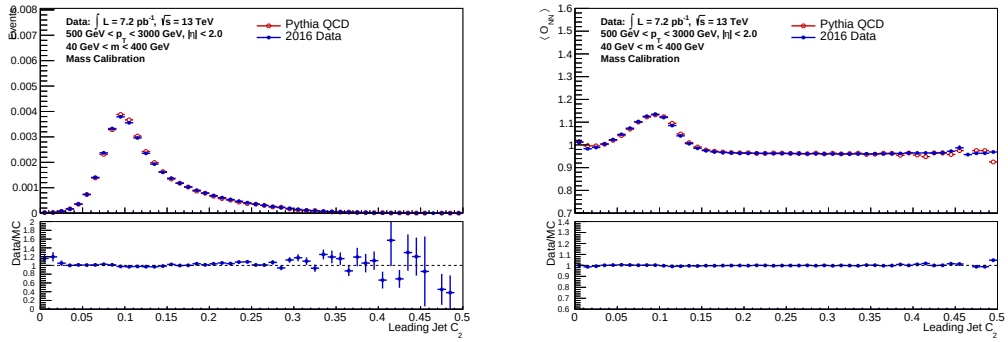



(b) $m$

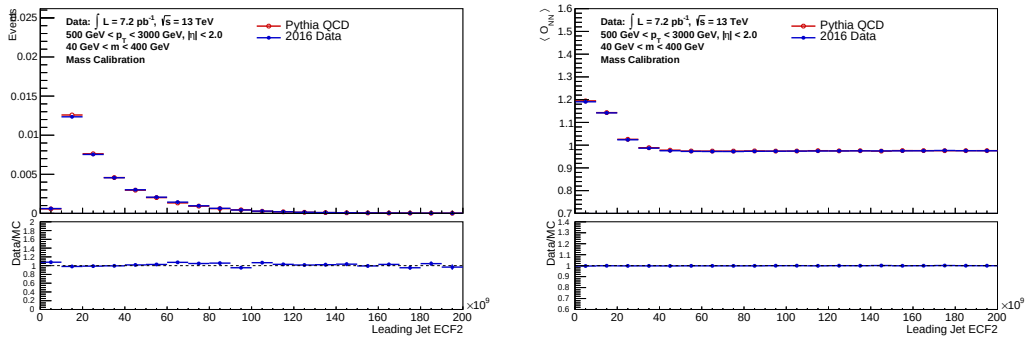



(c) $C_2^{track}$

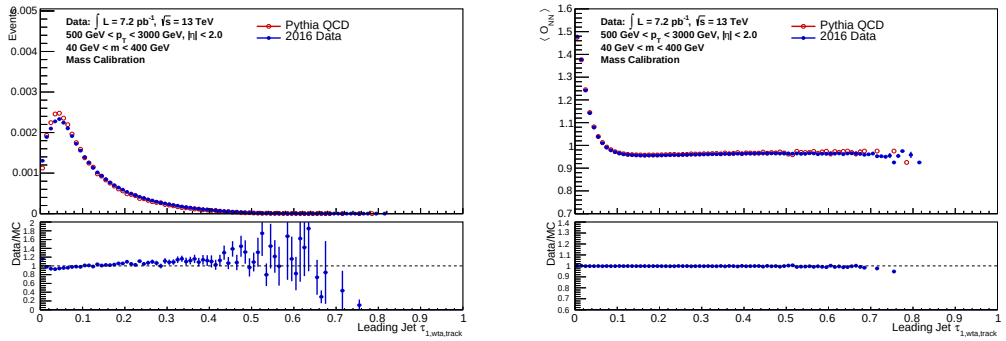
Figure 5.26: The plots on the left show a comparison of the $p_T/m$, $m$, and $C_2^{track}$ mass calibration network input variables in the data and MC samples. The right plots show the calibration network output as a function of these variables. The data/MC distributions show generally good agreement except for at low values of $C_2^{track}$ and $p_T/m$. Network output typically agrees between MC and simulation to $< 1\%$ with the largest divergence of $\sim 5\%$ for $C_2^{track} < 0.02$.




(a) $C_2$

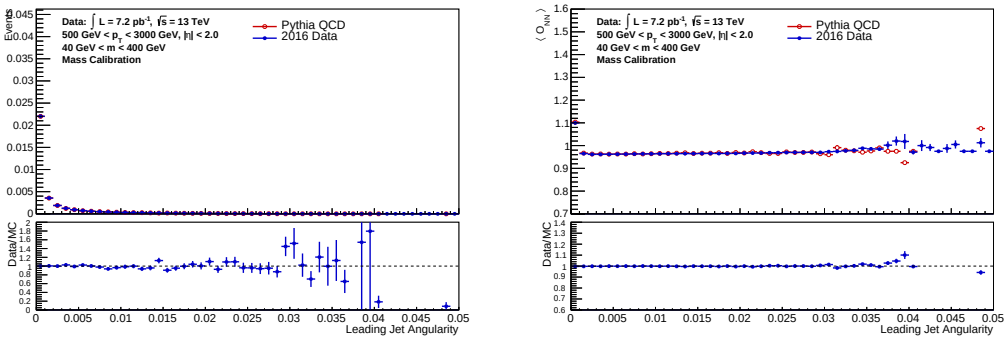



(b) $ECF_2$




(c) $\tau_{21,wta}^{track}$

Figure 5.27: The plots on the left show a comparison of the $C_2$, $ECF_2$, and $\tau_{21,wta}^{track}$ mass calibration network input variables in data and MC. The right-hand-side plots show the calibration network output as a function of these variables. The data/MC distributions show generally good agreement except for at low values of $C_2$. Network output agrees between MC and simulation to $< 1\%$




(a) $\tau_{21,wta}^{track}$

Figure 5.28: The plot on the left shows a comparison of the angularity mass calibration network input variable in data and MC. The right-hand-side plot shows the calibration network output as a function of these variables. Both input and network output distributions show good agreement.

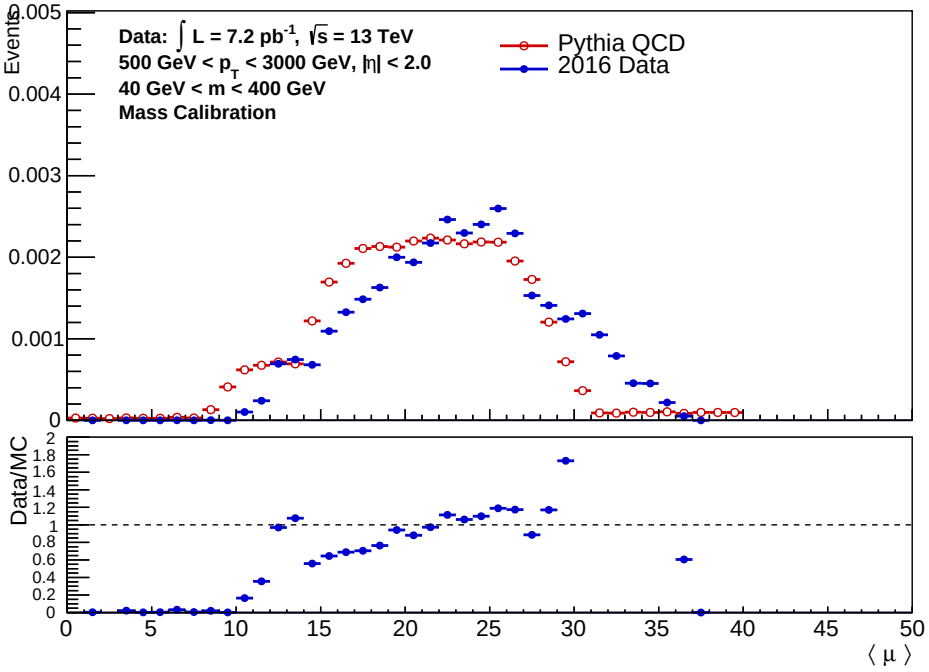
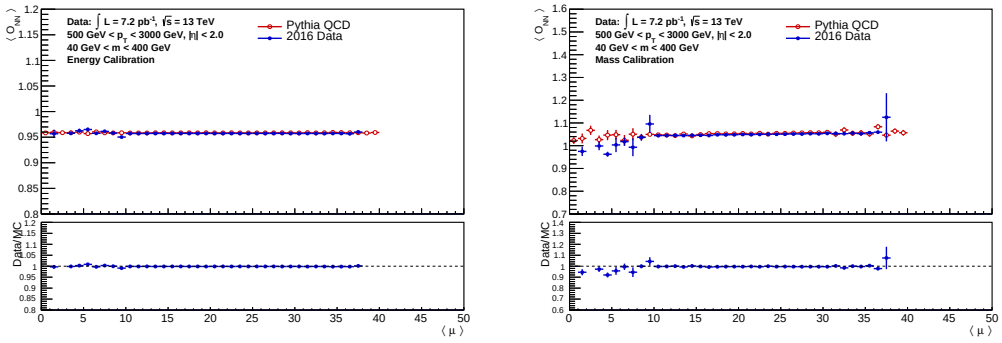



Figure 5.29: $\langle\mu\rangle$ distribution in the Pythia QCD and 2016 data samples.




(a) *Energy* (b) *Mass*

Figure 5.30: Mean energy and mass calibration network output as function of $\langle\mu\rangle$. Behaviour is similar between data and simulation and no significant dependence on $\langle\mu\rangle$ is observed.

inferior to, the current energy calibration. The lack of impact from additional variables or using larger networks suggests that our current calibration captures as much information about the jet energy response as is available using standard substructure variables. In addition, structures encoding the $p_T$ slices of the samples used for training are observed in the network output, causing non-physical structures in the energy and $p_T$ distributions when applied to data. The leaves the energy calibration network derived unsuitable for use in analysis. Both of these problems could be fixed, however, and this does not indicate improvement could not be made. For example, a convolutional approach using the jet constituents as inputs, similar to what has been done for top-tagging [127] could feasibly provide a better result provided sample-weighting errors were addressed.

Fortunately, the mass calibration proved more fruitful and showed that it is possible to improve on the current calibration using neural network methods. Using the IPCS algorithm a network was found which gave a 26% improvement to the resolution for both QCD and vector-boson-initiated jets while preserving or slightly increasing performance in all other datasets. Identifying the exact cause of this additional calibration power would be an interesting exercise, if it is coming from a single variable or small set of variables, then it should be feasible to train several networks with one or two variables removed and measure the change in performance. If it could be shown to be the result of information spread across multiple variables, then that would be a strong incentive to pursue a multivariate approach for future calibrations.

Both mass and energy networks were validated against data collected by ATLAS

during 2016 and their output was found to be consistent between data and MC to the 1% level in almost all regions studied. Despite the structures observed in the energy calibration, the networks behaved similarly for both data and MC indicating that the cross-variable correlations used are well-modelled in MC. The largest discrepancy observed was in the mass calibration network, where for low values of $C_2^{track}$ a difference in $\langle O_N N \rangle$ of 5% was found, though this was confined to an edge of the distribution. Combined with the validation performed on Herwig QCD samples, this strongly suggests that the mass network learnt general correlations rather than Pythia specific jet fragmentation/hadronisation or subtleties of the GEANT4 simulation used to model the detector for both Pythia and Herwig samples. The data/MC agreement for the energy network suggests that this is still true in the case of the energy network, the structures were the result of the network becoming biased towards classes of jet behaviour which occur on the edges of jet sample $p_T$ slices and were suppressed when sample and MC event weighting was applied to the Pythia sample.

It should be stressed that the data validation performed here was only sufficient to establish the (non-)existence of structural biases within the network output and is not a guarantee of a correct calibration on data. This study also did not consider uncertainties on the calibration results in any form. The use of tracking variables within the calibration makes the use of the double-ratio [89] method for measuring the uncertainties in data impossible; and while the jet energy scale, jet energy resolution and jet mass scale have been measured in-situ [128] this has not yet been shown for the jet mass resolution uncertainties. Any further extension to this study should focus on techniques for making these in-situ evaluations of the uncertainties.

On a final note, it would be interesting to apply the same techniques used in this study to variables which are not currently calibrated in analyses. The variable selection and network training procedure is extremely general and could easily be applied to $D_2$, $\tau_{21}$ or other similar jet substructure variables.

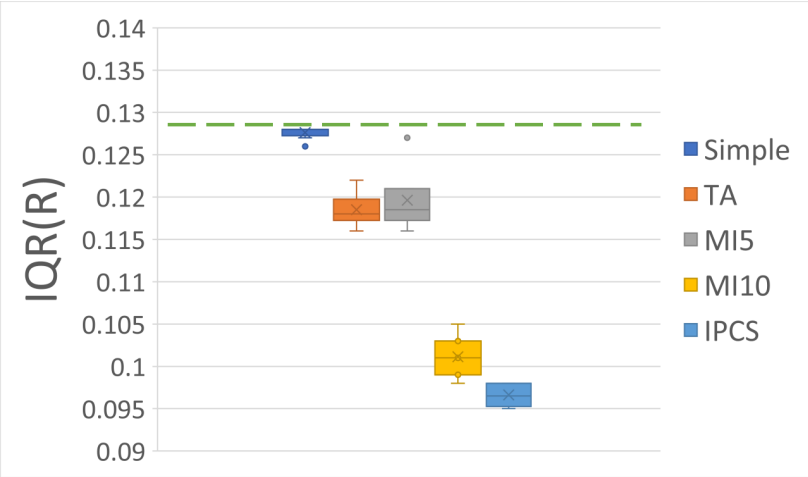



Figure 5.31: A box plot showing the median and interquartile range of the trained networks' $IQR(\mathcal{R})$ for each input set tried during the mass calibration. The green line shows the $IQR(\mathcal{R})$ of the existing calibration.

# Chapter 6

# Triggering on Large-Radius Jets

Even a perfect jet calibration is of no use if interesting jet events are missed by the trigger system. At the beginning of run-2, there were two approaches to triggering on large-radius jets within ATLAS, distinguished by their inputs to the jet-finding process. The first case was to use topoclusters, in a similar fashion to the offline jet reconstruction; the second was to run the jet-finding on pre-clustered and calibrated small-radius jets. This second method was initially favoured, as it seemed to provide many advantages such as providing calibrated jets without having to derive a separate online large-R jet calibration, and a significantly lower computation load due to the smaller number of inputs. This chapter begins with an evaluation of the efficiency of each type of trigger during ATLAS' 2015 run.

The increase in instantaneous luminosity provided by the LHC over the course of run-2 is causing additional challenges for the large-R jet trigger. When ungroomed, this type of jet is particularly sensitive to the effect of pileup due to its large area. Hence the jet trigger rate for a fixed energy threshold has increased far faster than the collaboration's capacity to store these collisions, forcing the energy threshold to increase. This was addressed in 2016 by the decision to focus the trigger on capturing decays from heavy particles and add jet trimming and a mass cut to the standard large-R jet trigger. Unfortunately, the further rate increases in 2017 and 2018 make even greater demands on the trigger. This chapter presents a feasibility study for an online ''anti-QCD'' tagger using jet substructure, and the possibility of using this to achieve a more sustainable trigger rate at current (or lower) energy thresholds.

| Chain Part | Description |
|---|---|
| a10 | Constructed from topo clusters using the anti-$k_T$, $R = 1.0$ algorithm. |
| a10r | Constructed from $R = 0.4$ anti-$k_T$ jets using the anti-$k_T$, $R = 1.0$ algorithm. |
| lcw | ''Local Cluster Weighted'' - Topoclusters used for input are reweighted from the EM to the hadronic energy scale. |
| sub | An area-based pileup subtraction is applied to each jet. |

Table 6.1: Some common parts of jet trigger chain names and their meanings.

### 6.0.1 Jet Trigger Chain Definitions

ATLAS trigger chain names are self-descriptive but somewhat unintelligible without definition. Since these chains names will be used throught this chapter a short introduction is in order. A standard jet trigger name can be broken down into four components; take, for example, HLT_j420_a10_lcw_sub_L1J100:

- HLT : This is a high level trigger algorithm (the alternative would be L1).
- j420 : Jet trigger with energy threshold of 420 GeV.
- a10 : Jets are constructed with the anti-$k_T$ algorithm with $R = 1.0$. Small-R anti-$k_T$ jets ($R = 0.4$) are the default and are used if this term is not present.
- lcw_sub : Any further jet properties; in this case, it lists applied calibrations.
- L1J100 : The L1 trigger which seeds a HLT chain.

All triggers considered by this chapter are HLT jet triggers seeded by an L1 jet trigger. The additional chain name parts encountered are described in table 6.1.

## 6.1 Large-Radius Jet Trigger Performance in 2015

As part of a larger paper on the 2015 ATLAS trigger performance [19], a study was made of the performance of the large-R jet trigger chains that were live during 2015. These chains were:

- HLT_j360_a10_lcw_sub_L1J100,
- HLT_j360_a10_sub_L1J100,
- HLT_j360_a10r_L1J100.

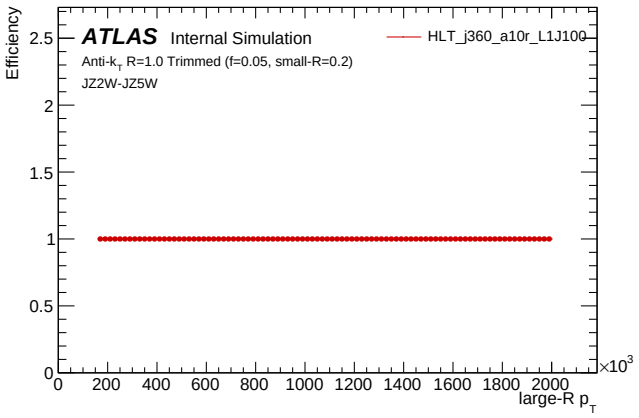



Figure 6.1: The efficiency of the trigger combination HLT_j110 + HLT_j360_a10r_L1J100 with respect to requiring just HLT_j360_a10r_L1J100, measured in QCD multijet MC.

The turn-on behaviour of each trigger was measured in data and evaluated as a function of various jet variables for QCD, W and top-like jet topologies. The behaviour of L1_J100 as a seed for large-R jet triggers was also examined.

### 6.1.1 Baseline

During the 2015 run, there was no lower threshold but prescaled large-R jet trigger available against which to measure the unprescaled large-R jet triggers. Instead, the efficiency was measured with respect to a set of lower threshold small-R triggers. The logical OR of the triggers HLT_j15, HLT_j25, HLT_j35, HLT_j55, HLT_j85, and HLT_j110, was used, with HLT_j110 chosen as the upper limit to preserve 100% efficiency with respect to the 360 GeV large-R jet trigger. The OR of the triggers is used to compensate for the loss of statistics due to trigger prescales. Figure 6.1 shows that in MC, in which prescales were not applied, all events which passed the HLT_j360_a10r_L1J100 trigger also passed the HLT_j110 trigger.

The full dataset collected during the 2015 *pp* run was used for the efficiency measurement, events were considered if they met the following preselection requirements:

- Pass the good runs list selection. We only consider lumi-blocks from runs where the detector state was classified as "All-Good", defined by the GRL: data15_13TeV.periodAllYear_DetStatus-v73-pro19-08_DQDefects-00-01-02

_PHYS_StandardGRL_All_Good.xml

- Pass event cleaning cuts, which remove events that were incomplete or with bad LAr or Tile data quality [119].
- $\geq 1$ large-R and $\geq 1$ small-R offline jets with $p_T$ greater than 200 GeV and 20 GeV respectively.
- Pass jet cleaning cuts, removing events with a small-R jet with $p_T > 20$ GeV labelled ''BadLoose'' [116].

When measuring the L1 trigger efficiency, the baseline dataset was *all* 2015 data passing this selection.

### 6.1.2 Efficiency Measurements

The trigger efficiency was measured as a function of the offline leading jet $p_T$ (figure 6.2) and it was found that the 99.95% efficiency plateaus were 380 GeV, 480 GeV, and 500 GeV for HLT_j360_a10_lcw_sub_L1J100, HLT_j360_a10_sub_L1J100, and HLT_j360_a10r_L1J100 respectively. However, full efficiency is not reached due to losses in the level 1 trigger, which continue to around 700 GeV. The L1 jet trigger is a sliding window algorithm applied to the sum of calorimeter tower energies in $\eta \times \phi$-space that searches for energy maxima in $0.2 \times 0.2$ windows. Hence, it is not unexpected that a small number of jets with evenly distributed energy would fail this cut.

The poor performance of the reclustered trigger was unexpected however as previous analysis specific measurements had placed its turn-on plateau much lower (see §7).

The impact of several other variables was investigated to see whether any were strongly correlated with each trigger's efficiency, the jet $\eta, \phi, m_J$ and $D_2$ were considered. Figure 6.3 shows that a strong dependence on the jet mass and $D_2$ variables were observed for all large-R trigger algorithms.

Going further, the jet sample was divided into three regions, W/Z tagged, top tagged, and no tag, by running the 2015 ATLAS W/Z and top tagging algorithms on the offline jets matched to the trigger jets considered. The trigger efficiency was then measured for each sample (figure 6.5). Here it was found that for data passing the W/Z tag, the performance of the reclustered trigger significantly improved, no slow plateau is observed, and its performance is equal to or better than the lcw-calibrated non-reclustered trigger. This matches both the previously seen dependence on $D_2$

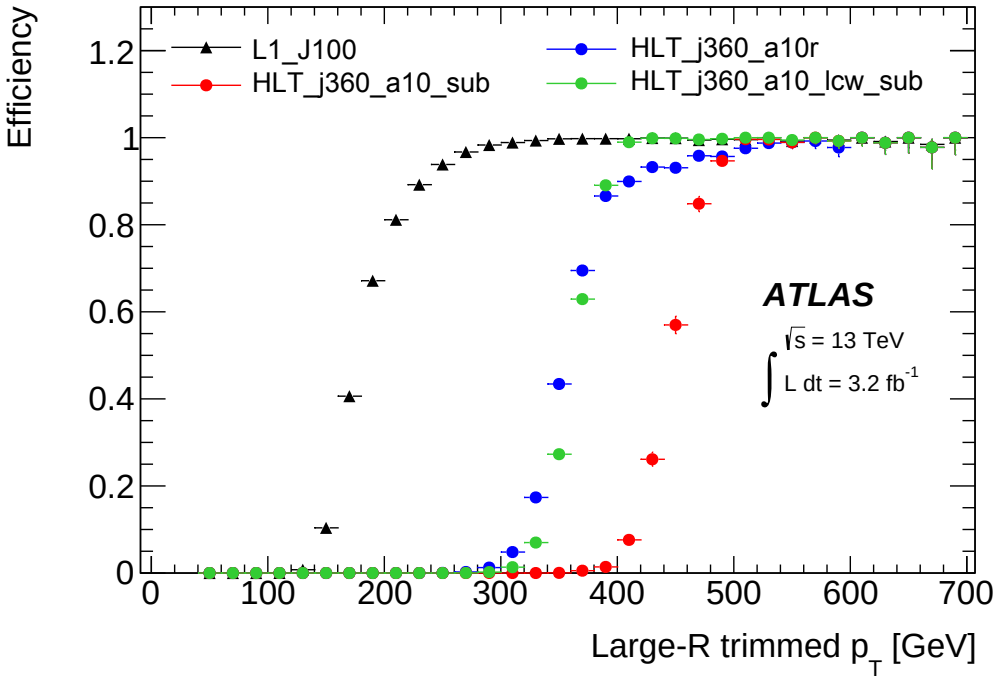
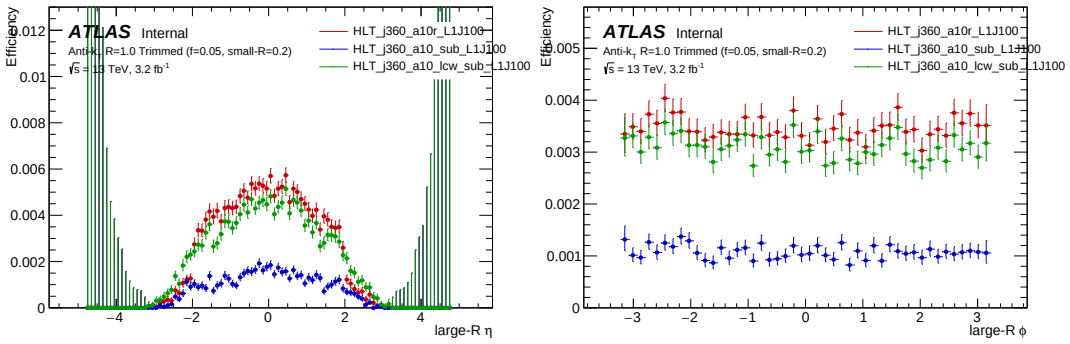


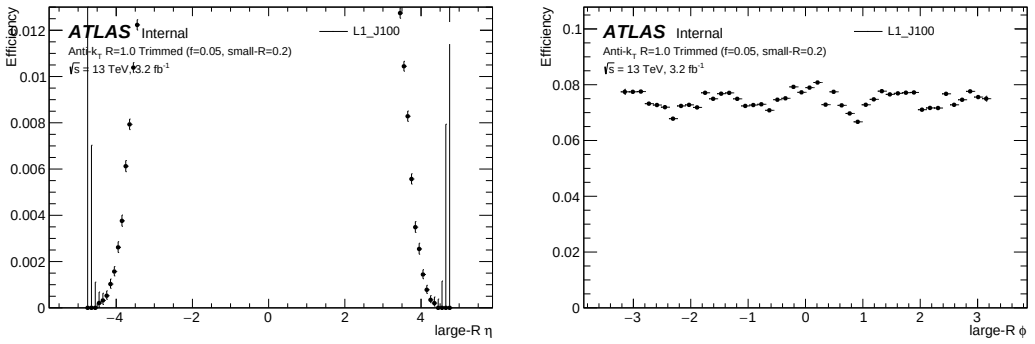

Figure 6.2: The turn-on of the L1_J100 trigger and each large-R trigger as a function of leading jet $p_T$. Note that the reclustered jet begins to turn on slightly faster than the lcw-scale non-reclustered trigger but is much slower to reach the efficiency plateau.



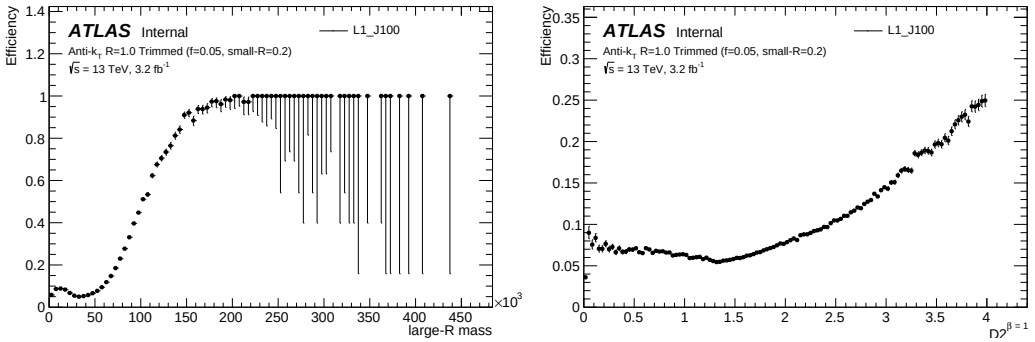

(a) $\eta$ (b) $\phi$

(c) $m_J$ (d) $D_2$

Figure 6.3: The turn on of each large-R trigger as a function of the leading jet $\eta$ (a), $\phi$ (b), mass (c), and $D_2$ (d), strong biases in the performance are seen in both mass and $D_2$ but the acceptance is relatively flat in $\phi$ and $\eta$ for $|\eta| < 2.0$.

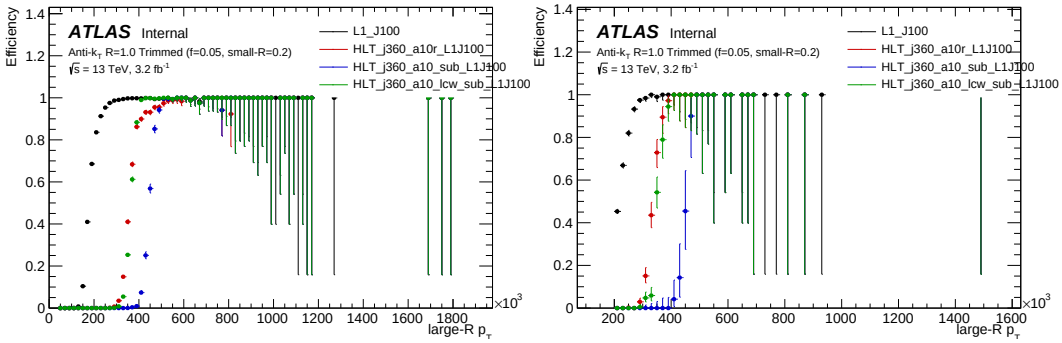



(a) $\eta$ (b) $\phi$



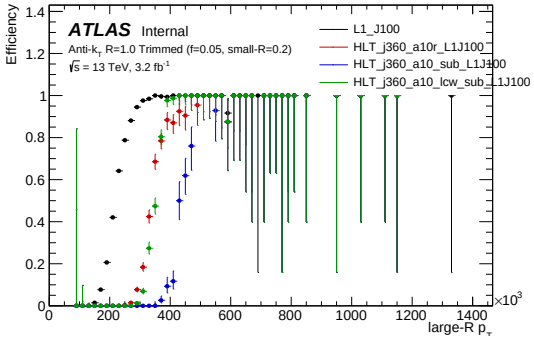

(c) $m_J$ (d) $D_2$

Figure 6.4: The turn on of L1_J100 as a function of the leading jet $\eta$ (a), $\phi$ (b), mass (c), and $D_2$ (d). Again, strong biases in the performance are seen in both mass and $D_2$ but the acceptance is relatively flat in $\phi$ and $\eta$ for $|\eta| < 2.0$.




(a) No Tag

(b) W/Z Tag




(c) Top Tag

Figure 6.5: The turn-on of the L1_J100 trigger and each large-R trigger as a function of leading jet $p_T$ for untagged (a), W/Z-tagged (b) and top-tagged (c) data samples.

and the $X \to VV \to JJ$ analysis' trigger measurements. The large effect is likely because a low $p_T$ QCD jet's energy is typically diffuse around the jet's area, and the reclustering process is not likely to pick up all of the small-R jets containing the offline large-R jets energy. However, low-$p_T$ hadronic W/Z decays will usually be fully captured in two small-R jets which are combined during the reclustering process.

### 6.1.3 Conclusion

The fully-calibrated topocluster-based trigger chain HLT_j360_a10_lcw_sub_L1J100 was clearly the best performing of the three high-level trigger chains examined, obtaining 100% efficiency at $p_{T,J} \approx 380$ GeV, well ahead of the other two triggers. For QCD-like jets the reclustered trigger clearly performed less well, despite a similar initial turn-on, however it performed nearly identically to the topocluster trigger for jets passing a W/Z boson selection. The EM-scale topocluster trigger,

HLT_j360_a10_sub_L1J100 performed worse than the other two triggers for all samples.

Some inefficiencies at high $p_T$ were observed in all high level triggers, these also appear to be present in the L1 trigger so are likely the result of a failure to run the chain at all when certain jet topologies occur. These inefficiencies are not seen for W/Z or top tagged jets so are not likely to effect any analyses making use of the triggers.

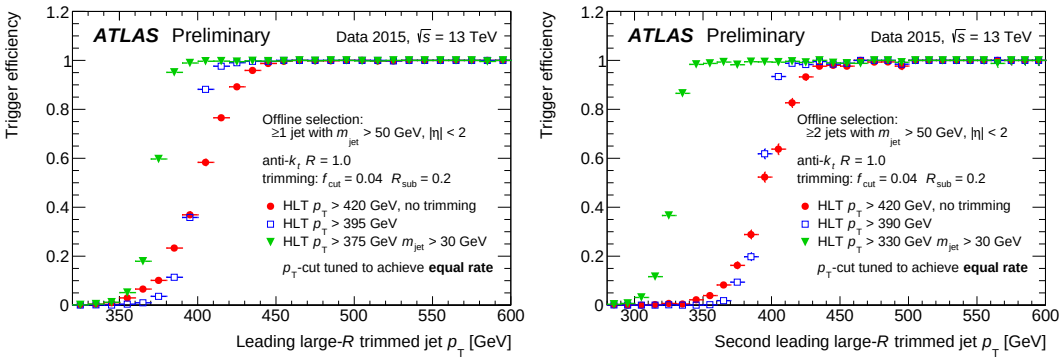
## 6.2 An "Anti-QCD" Jet Trigger

As the instantaneous luminosity received by ATLAS has increased, the requirements placed on the trigger system have become substantially tighter. The rising data rate means that trigger thresholds must increase to keep the trigger rate below 1kHz. Unfortunately, this can have a direct impact on analyses which no longer have access to events with jets below the increased trigger threshold. This has been a significant problem for the large-R jet trigger, which, due to the large area of the jets, is particularly susceptible to increases in pileup. The higher rate forced the minimum trigger threshold from 360 GeV in 2015 to 420 GeV in 2016 and, unmodified, would have forced the trigger substantially higher in 2017.

A saving grace is that the large-R jet trigger is primarily used for collecting hadronically-decaying heavy objects, i.e. the W/Z/H/top, but quark and gluon jets dominate its rate. Since identification of these objects has been well studied offline, it was possible to apply similar restrictions at trigger level to reduce the rate without affecting analyses making use of the trigger [129].

The 2017 trigger has been designed to make use of two offline techniques, jet trimming, reducing the pileup contributions to the jet energy and mass, and a minimum requirement on the jet mass. Both of these are applied more loosely than their offline equivalents, the trimming has a 4% threshold rather than 5%, and the mass cut is placed at 30 GeV, but this still substantially reduced the rate. The single large-R jet trigger threshold was reduced to 395 GeV, but the total rate was maintained.

Additionally, applying these selection criteria makes it useful to consider dijet triggers. These were not especially helpful for the simple jet trigger, as most events with a single hard jet will possess a second with a similar energy scale. In simulation applying a 2nd jet requirement would typically reduce the rate by around 10%. However, the mass of the two jets should be relatively uncorrelated, and it




(a) (b)

Figure 6.6: A comparison of the ''equal rate'' turn-on curves for single jet (a) and dijet (b) triggers. The red curve is the 2016 large-R jet trigger, the blue curve a large-R trigger using trimmed jets, and the green curve is a trigger requiring trimmed jets with a mass greater than 30 GeV. The threshold for each type of trigger is chosen to give a total rate equal to that of the 2016 trigger. Figures are public plots from [129].

was found that a significant fraction of events could be removed by requiring at least two jets to pass the trigger requirements. For 2017 there is a specialised large-R dijet trigger with a minimum threshold of 330 GeV.

Unfortunately, from a trigger perspective at least, the LHC and ATLAS do not stand still and are planning on delivering further increases to the event rate in 2018, which would again require significant increases to the trigger thresholds, even with the new trimming and mass requirements. We performed a proof-of-concept study into the possibility of applying further substructure selection in the jet trigger to try and maintain (or even lower) the current thresholds.

### 6.2.1 Datasets

Three datasets were used in this study to measure the efficiency for QCD, vector boson and top jets. Data passing the jet trigger represents a relatively pure sample of QCD jets but MC was required for the $W/Z$ and top jets. The samples used were:

- Data: $1.6 fb^{-1}$ of data collected during period J of the 2015 data-taking.
- Vector Boson: A Pythia8 $W' \to WZ \to qqqq$ sample with $m_{W'} = 1$ TeV.
- Top: A Pythia8 $Z' \to t\bar{t}$ sample with hadronically decaying tops. $m_{Z'} = 1$

TeV.

Both MC samples used the NNPDF23LO PDF set [114] and A14 tune [113]. For the data sample, events are required to pass a GRL requirement and data quality cuts. For all samples, at least 1 offline large-R jet with $p_T > 300$ GeV, $m > 50$ GeV and $|\eta| < 2.0$ is required and 1 online jet with $|\eta| < 2.8$.

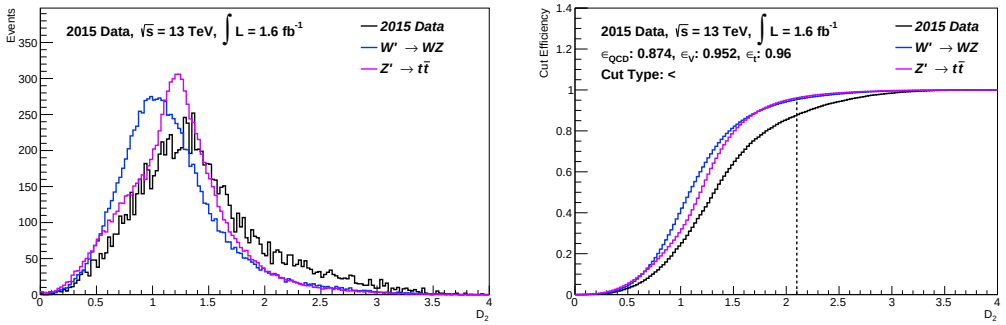
### 6.2.2 QCD-Tagging

The large-R jet trigger is used for a variety of physics objects with substantial differences in their internal structure, e.g. a 2-prong W decay vs a 3-prong top decay. As such, what is required for a trigger used for both is an ''Anti-QCD'' tagger, which is quite a different problem than those solved by the existing taggers. The selection is further constrained due to being run at trigger level. Firstly, the signal efficiency must remain as high as possible as any events missed by the trigger are gone permanently. Secondly, it must be optimised for good background rejection at a 95% signal efficiency rather than the 50% or even 25% efficiency of the offline taggers. Finally, it must be simple enough to run on the trigger without substantially slowing down the existing jet trigger algorithms.

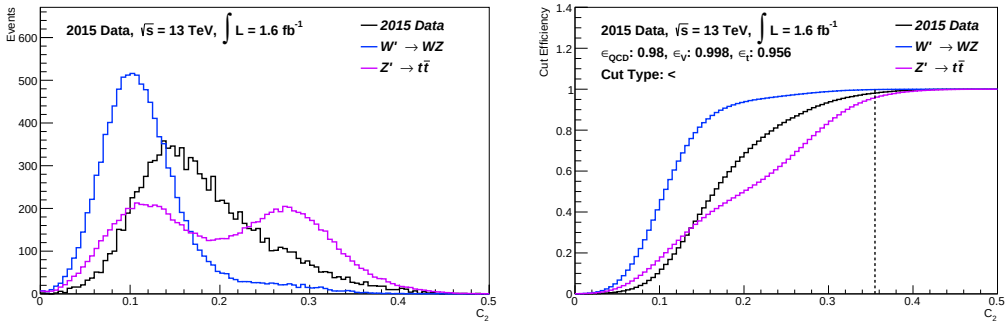
The initial approach was to try and find an optimal 1D or 2D cut to discriminate against QCD jets while maintaining a 95% signal efficiency for W, Z and top jets with respect to the existing trigger selection. It is expected that the selection efficiency for Higgs jets should be similar to that of Ws and Zs.

A wide range of variables, both existing and new, were investigated. The potential of each variable was estimated by finding the cut which maximised QCD rejection in data, at a fixed acceptance working-point for both the WZ and top MC samples. Interestingly, the performance of variables already used for tagging varied wildly. The $D_2$ and $\tau_{21,wta}$ variables showed reasonable discriminatory power while preserving both top and WZ samples, but the similar $C_2$ and $\tau_{32,wta}$ variables were only useful for discriminating WZ and top jets separately, as shown in figures 6.76.8.

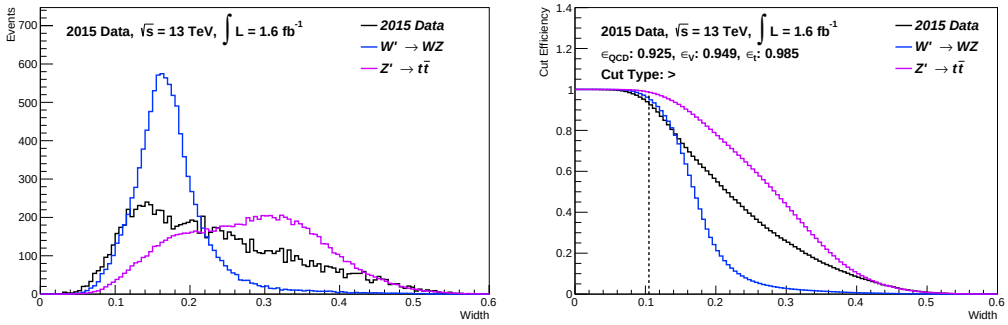
A set of new or less common variables combinations, designed explicitly for anti-




(a) $D_2$

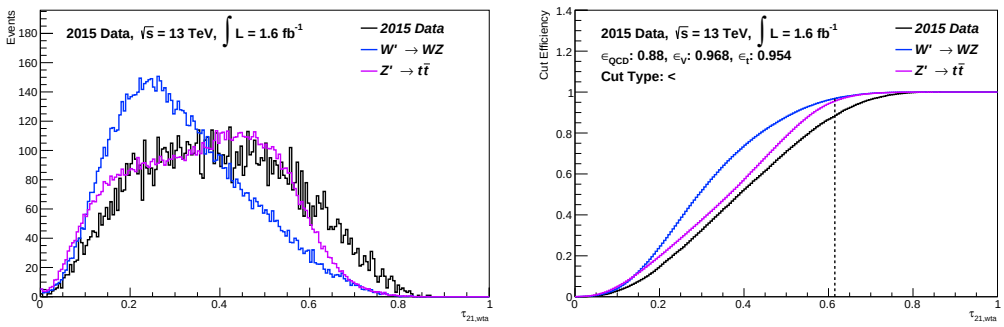



(b) $C_2$

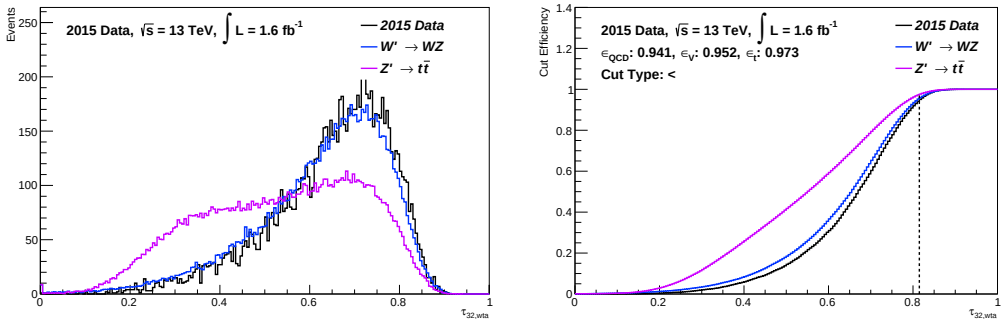



(c) Width

Figure 6.7: Each subfigure shows on the left: the distributions of the $D_2$, $C_2$, and Width variables for 2015 data, $W' \to WZ$, and $Z' \to t\bar{t}$ MC samples, and on the right: the percent of events in each sample which pass a cut corresponding to the x-axis value. This cut is of the form $v_{jet} < v_{cut}$ for all figures except (c) which requires $v_{jet} > v_{cut}$. The dashed line marks the cut with maximal QCD rejection while maintaining at least 95% signal efficiency for the WZ and $t\bar{t}$ samples.




(a) $\tau_{21,wta}$




(b) $\tau_{32,wta}$

Figure 6.8: Each subfigure shows on the left: the distributions of the $\tau_{21,wta}$ and $\tau_{32,wta}$ variables for 2015 data, $W' \to WZ$, and $Z' \to t\bar{t}$ MC samples, and on the right: the percent of events in each sample which pass a cut corresponding to the x-axis value. This cut is of the form $v_{jet} < v_{cut}$ for all figures. The dashed line marks the cut with maximal QCD rejection while maintaing at least 95% signal efficiency for the WZ and $t\bar{t}$ samples.

| Variable | Cut Type | Value | QCD Rejection [%] | $\epsilon_V$ [%] | $\epsilon_t$ [%] |
|---|---|---|---|---|---|
| $D_2$ | $<$ | 2.1 | 12.6% | 95.2% | 94.3% |
| $C_2$ | $<$ | 3.6 | 2.0% | 99.8% | 95.6% |
| Width | $>$ | 7.5 | 9.2% | 94.9% | 98.5% |
| $\tau_{21,wta}$ | $<$ | 0.62 | 12.0% | 96.8% | 95.4% |
| $\tau_{32,wta}$ | $<$ | 0.82 | 5.9% | 95.2% | 97.3% |
| $\tau_{31,wta}$ | $<$ | 0.39 | 16.3% | 95.3% | 96.3% |
| $\tau_{31,wta}\tau_{21,wta}$ | $<$ | 0.22 | 16.6% | 95.2% | 95.4% |
| $\tau_{321,wta}$ | $<$ | 0.96 | 16.7% | 95.1% | 95.0% |

Table 6.2: The cut which gives maximum QCD rejection while maintaining at least 95% signal efficiency for both WZ and $t\bar{t}$ samples for each investigated variable.

QCD tagging were examined to try and gain further discriminatory power.

$$\tau_{31,wta} = \frac{\tau_{3,wta}}{\tau_{1,wta}} \tag{6.1}$$

$$\tau_{31,wta}\tau_{21} = \frac{\tau_{3,wta}\tau_{2,wta}}{\tau_{1,wta}^2} \tag{6.2}$$

$$\tau_{321,wta} = \frac{\tau_{3,wta} + \tau_{2,wta}}{\tau_{1,wta}} \tag{6.3}$$

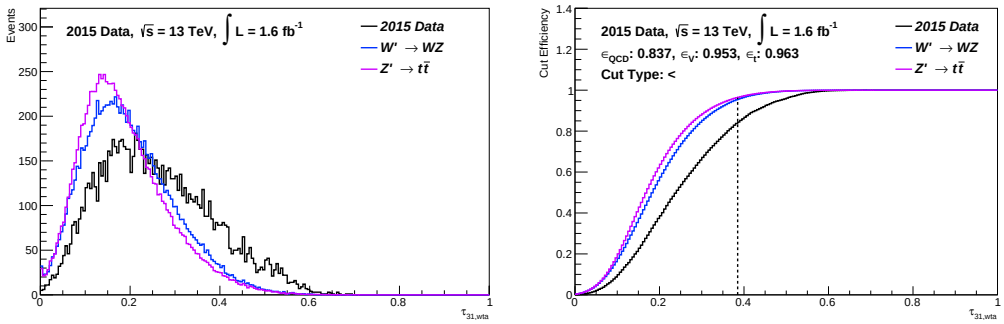
The distributions of these variables are shown in figure 6.9. All three of these new variables outperformed all currently used tagging variables investigated during this study, rejecting around 16% of QCD jets while maintaining a 95% signal efficiency. The optimal cuts for each variable are shown in table 6.2

#### 2D Selections

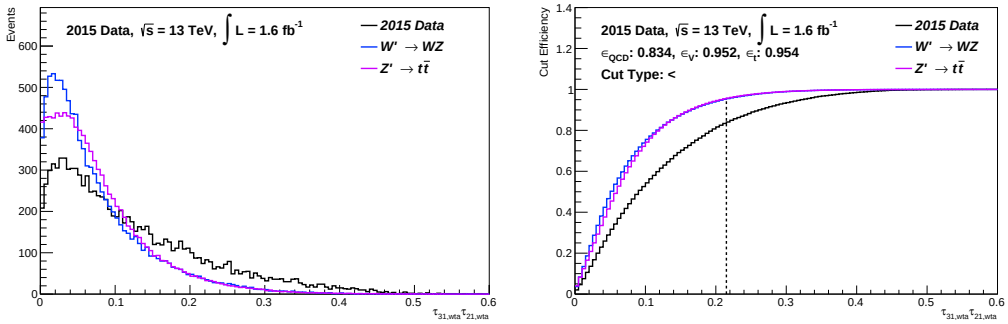
Linear 2D cuts on the same variables were investigated optimised alongside the simpler 1D cuts but were not found to give any significant improvement over the single anti-QCD variables.

### 6.2.3 Rates

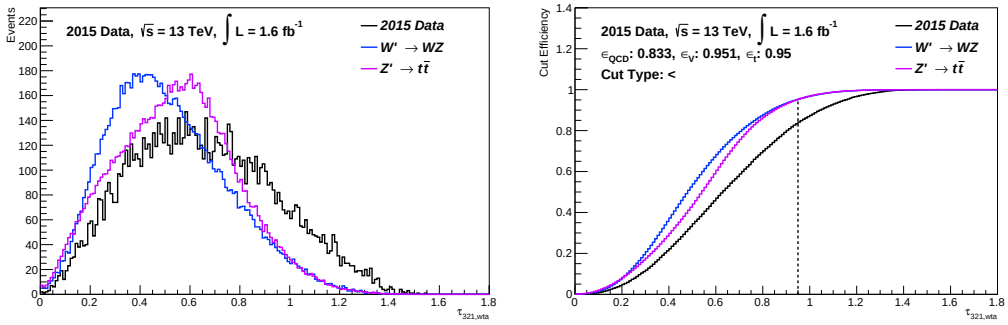
After variable optimisation, the expected trigger rates as a function of the trigger threshold were calculated for each anti-QCD variable cut, with and without the 30 GeV mass cut. The estimated rates as a function of trigger threshold are shown in figure 6.10 and table 6.3 shows the energy threshold for each cut that would give a




(a) $\tau_{31,wta}$

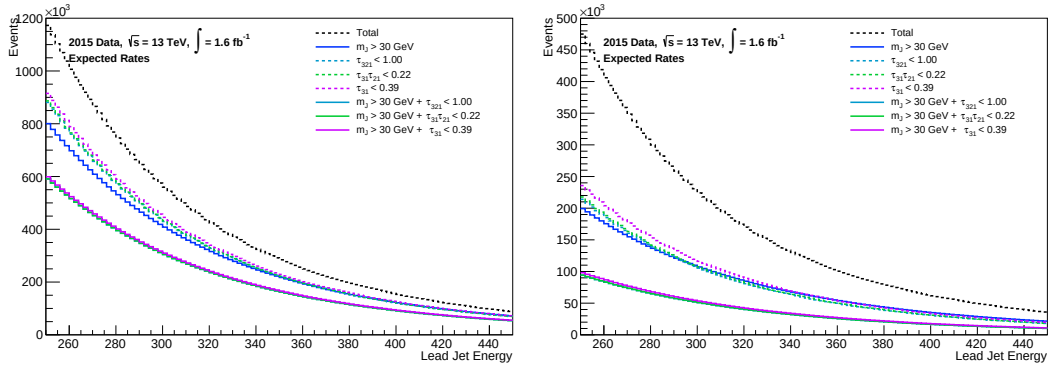



(b) $\tau_{31,wta}\tau_{21,wta}$




(c) $\tau_{321,wta}$

Figure 6.9: The observed distributions and cut efficiencies for the three anti-QCD tag variables, $\tau_{31,wta}$, $\tau_{31,wta}\tau_{21,wta}$, and $\tau_{321,wta}$. All three significantly out-perform the other variables considered with nearly indistinguishable rejection power.




(a) Single Jet (b) DiJet

Figure 6.10: The estimated trigger rate as a function of the leading jet energy after the mass and substructure cuts have been applied for a single (a) and dijet (b) trigger.

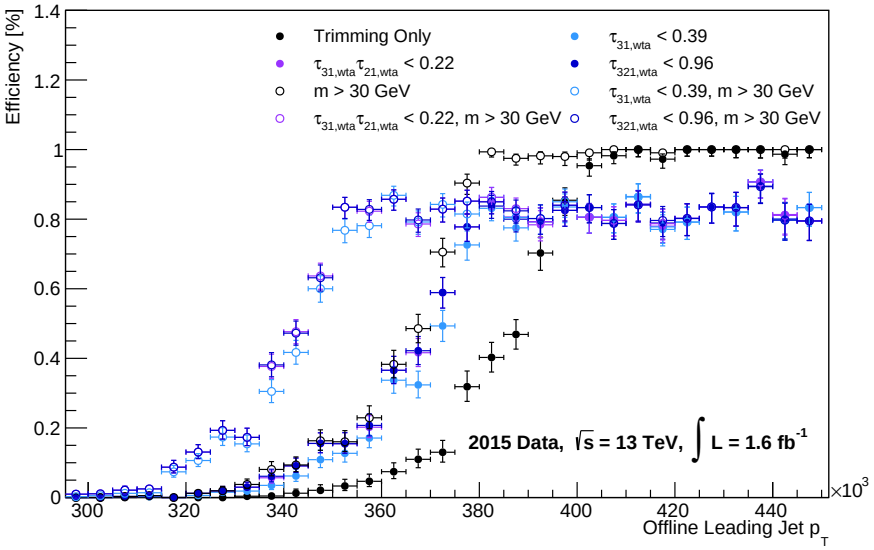
total trigger rate equal to that of the 420 GeV trigger run in 2016 for both single and dijet triggers.

The application of anti-QCD tagging at the trigger level gives a substantial reduction in trigger rates beyond that provided by trimming and applying a mass cut. All else being equal, it would allow the large-R trigger threshold to be reduced significantly, to around 350 GeV in the case of the single jet trigger. Since the substructure information is uncorrelated between jets in an event, the dijet trigger obtains an even greater reduction in rate. However, the proposed thresholds should be taken with a pinch of salt as they fall into the region where the level 1 jet trigger is not fully efficient for large-R jets. They also assume that the operational jet trigger rate is left unchanged. The emulated turn-on curve for these equal-rate single jet triggers is shown in figure 6.11.

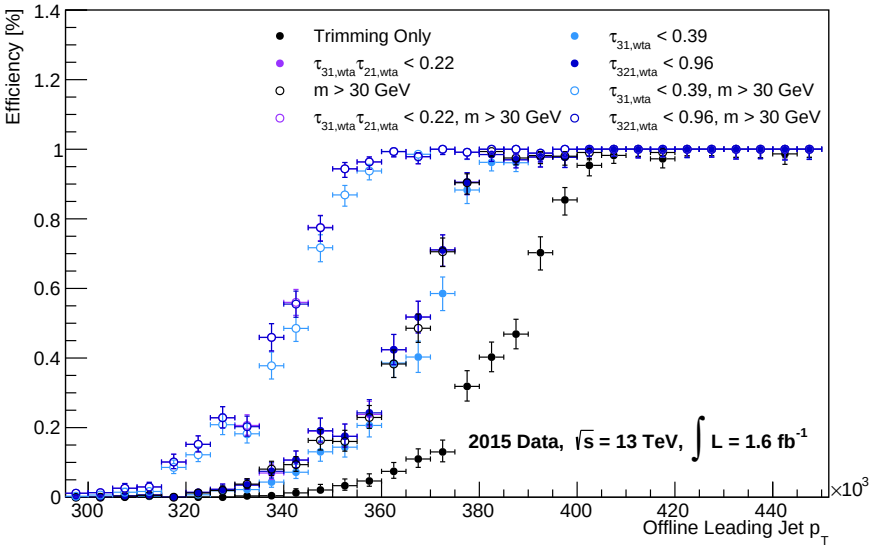
## 6.3 Conclusion

The two studies presented here represent a small part of the evolution in thinking within ATLAS about triggering on large-R jets.

The performance of the three large-R jet chains run by the ATLAS trigger during 2015 was found to vary significantly both between chains and between jet topologies. The HLT_j360_a10_lcw_sub_L1J100 trigger which constructed online jets from local-cluster weighted topoclusters clearly outperformed both the reclustered




(a)




(b)

Figure 6.11: A comparison of the "equal rate" turn-on curves as a function of calibrated offline leading jet $p_T$ for emulated single jet triggers with the the thresholds given in table 6.3. (a) shows the efficiency relative to events whose offline leading jet has $m > 50$ GeV and therefore the substructure triggers will not reach 100% efficiency, due to their rejection of jets failing the substructure cut. (b) shows the trigger efficiency for events where the leading (online) jet also passes this substructure cut. Note that the $\tau_{31,wta}\tau_{21,wta}$ curve is shown but is generally covered by the $\tau_{321,wta}$ curve.

| Cut | Single Jet Threshold [GeV] | Dijet Threshold [GeV] |
|---|---|---|
| Trimming only | 395 | 395 |
| $m > 30$ GeV | 374 | 342 |
| $\tau_{31,wta} < 0.39$ | 377 | 344 |
| $\tau_{31,wta}\tau_{21,wta} < 0.22$ | 374 | 337 |
| $\tau_{321,wta} < 0.96$ | 374 | 336 |
| $m > 30$ GeV + $\tau_{31,wta} < 0.39$ | 351 | 282 |
| $m > 30$ GeV + $\tau_{31,wta}\tau_{21,wta} < 0.22$ | 349 | 279 |
| $m > 30$ GeV + $\tau_{321,wta} < 0.96$ | 349 | 279 |

Table 6.3: The ''equal rate'' thresholds for each trigger selection relative to the untrimmed HLT_j420_a10_lcw_sub_L1J100 trigger. The reduction in rate is similar for all anti-QCD tag variables and seems to be orthogonal with the reduction obtained by applying a mass cut.

and the EM-scale topocluster jet triggers; achieving full efficiency at 380 GeV. The reclustered trigger, HLT_j360_a10r_L1J100 saw the largest performance differences based on jet topology. For W/Z tagged jets its performance was similar to that of the HLT_j360_a10_lcw_sub_L1J100 trigger, achieving full efficiency at 380 GeV, but it took until 450 for this to occur in untagged jets. The EM-scale trigger, HLT_j360_a10_sub_L1J100 performed badly in all topologies, also achieving full efficiency at around 450-500 GeV. The result formed a small part of a paper covering the a broad range of ATLAS trigger performance topics in 2015 [19].

During 2016 and 2017, the type of triggers which were evaluated by this study have become too accepting for use in the current high-pileup environment of the LHC and thresholds would be pushed unacceptably high if the current trigger rate was maintained. The currently operational solution to this is to apply jet trimming and a loose mass cut to online jets, taking advantage of large-R jet's primary use as proxies for hadronic heavy object decays. This line of thinking was taken further in the second study presented. Here we showed that combinations of the N-subjettiness variables can be used online to suppress QCD jets in favour of both W/Z and top decays. Using these, the trigger threshold could be reduced without any increase in rate when compared to the current trimmed jet triggers. No advantage was observed in combing any pair of these variables over the presented single variable cuts. Interestingly, shortly before the submission of this thesis a paper was published showing a very similar approach to anti-QCD tagging making use of the N-subjettiness vari-

ables combined using a neural network [130]. The study was not strictly comparable as it was entirely in simulation and restricted to offline, rather than online jets but no additional QCD rejection appears to have been achieved by this approach when requiring 95% signal acceptance.

Only a small amount of effort is needed to integrate the anti-QCD variables proposed by this study into the current trigger algorithms and this is currently under-way within the ATLAS collaboration.

# Chapter 7

# Search for Heavy Resonances with Bosonic Jets

There is a common belief that the problems in the Standard Model highlighted in §3.4 will be solved by the introduction of new particles, forces, or interactions, at a higher energy scale (or with smaller cross-section) than those currently known. It is further *hoped* that we will be able to observe these processes at the LHC.

If there were just one, or a small number of, known model(s) that could address these issues[1] then it would greatly simplify the search for new physics, but, unfortunately, there seems to be an infinite space of possible extensions to the Standard Model. Therefore, it is useful to carry out model-independent searches and constrain the parameter space of many such models at once.

The following chapter discusses a search for new TeV-scale particles decaying to the electroweak bosons. This is a common component of BSM models and, aside from the graviton [60, 62, 64, 131] and HVT models [65] discussed in §3.4, it is also a signature of Grand Unified Theories (GUTs) [132–134], Higgs doublet models [135], Technicolour [136, 137], and other generic composite Higgs models [138].

This analysis targets the fully-hadronic decay mode of this interaction, where both vector bosons decay to pairs of quarks (the $llqq$, $l\nu qq$ and $\nu\nu qq$ decay modes have also been studied at ATLAS and CMS [54, 139–142]). When a TeV scale particle decays to vector bosons they will be produced with a large Lorentz boost and the resulting final-state hadrons will be highly collimated in the detector, with a signif-

[1] Prior to the operation of the LHC, supersymmetry was widely believed to be such a model. So far this belief has not been matched with any actual evidence.

icant overlap in the showers from each quark pair. These are reconstructed as two back-to-back large-R jets. The jet grooming and substructure techniques from §4.3 can be used to identify boson-initiated jets and suppress the QCD background. After event selection has been applied, we conduct the search by looking for a resonant structure on the smoothly falling dijet mass distribution.

This chapter presents a search for these diboson resonances using the full $\sqrt{s} = 13$ TeV dataset collected by the ATLAS detector in 2015 and 2016, corresponding to a combined integrated luminosity of $36.7 fb^{-1}$. Full details are given in an ATLAS internal note [143], and a public paper summarising this search was published in 2018 [11].

## 7.1 Previous Searches

Searches using boosted bosons in the fully hadronic final state were performed for the first time during the LHC's first run, at both ATLAS [17] and CMS [144]. The ATLAS analysis found a $2.5\sigma$ excess in the region of 2 TeV, which was not excluded by the CMS analysis. This result (as is typical) led to some excitement in the theory community and many papers were written detailing models that attempted to explain the excess (see figure 7.1). Searches in the semi-leptonic [139–141] (where only one vector boson decays hadronically) and fully-leptonic [142] channels were also performed during run-1 and ATLAS released a further study combining the limits from the hadronic, semi-leptonic and fully-leptonic channels, but the leptonic searches observed no further excess and the overall significance was greatly reduced [54]. The strongest limits set with run-1 data are from the ATLAS combined result, which excludes $m_{W'} < 1.8$ TeV and $m_{G_{RS}} < 0.8$ TeV (for a bulk graviton with $k/\bar{M}_{planck} = 1$) at a 95% confidence level. Figure 7.2 shows some of the limits set by the hadronic and combined studies.

Following the restart, both ATLAS and CMS collaborations released searches using the 2015 dataset, taking advantage of the increase in sensitivity due to the higher centre of mass energy [13, 57]. The CMS search combined the $l\nu qq$ and $qqqq$ search channels and the ATLAS search used the $llqq, l\nu qq, \nu\nu qq$ and $qqqq$ channels. Despite the much lower integrated luminosity (3.2 $fb^{-1}$ and 2.6 $fb^{-1}$ for ATLAS and CMS respectively) than the run-1 analyses, these limits (shown in figure 7.3) were comparable to, or better than, the previous limits. CMS placed a 95% exclusion limit for $m_{W'} < 2.0$ TeV for a HVT Model B $W'$ and placed limits on the bulk

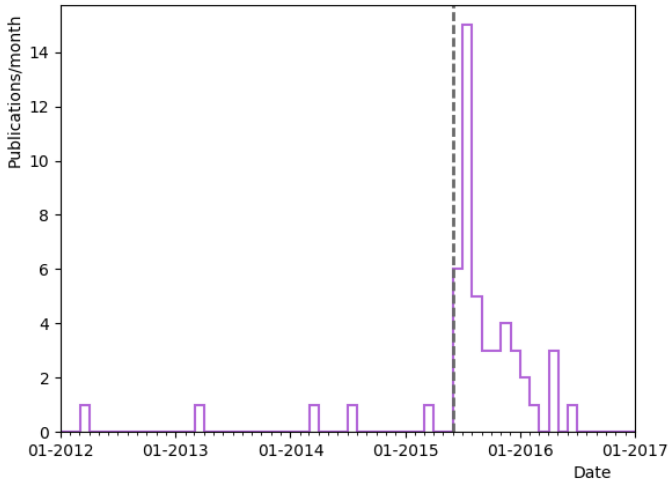


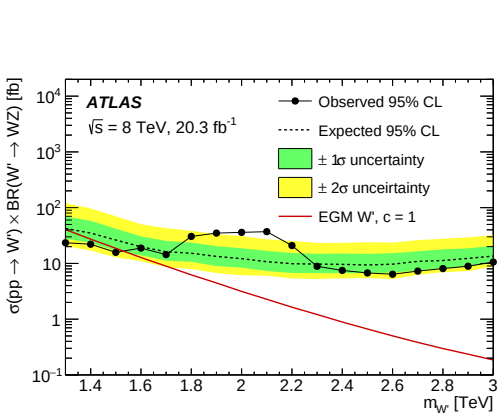

Figure 7.1: The number of papers published to the ‘‘hep-ph’’ or ‘‘hep-th’’ sections of arxiv each month whose title matches the regular expression ‘‘(diboson|WW|WZ|ZZ|VV).*excess’’. The grey line marks the publication of the ATLAS fully-hadronic diboson search on the 11$^{\text{th}}$ November 2015.

RS-graviton cross-section for $k/\bar{M}_{planck} = 0.5$. ATLAS placed a limit of $m_{W'} > 2.6$ (2.4) TeV on the HVT Model B (Model A) $W'$ and $m_G > 1.1$ TeV for a bulk RS graviton with $k/M_{planck} = 1.0$. A further limit of $M_X > 2.65$ TeV was set on a generic heavy scalar.

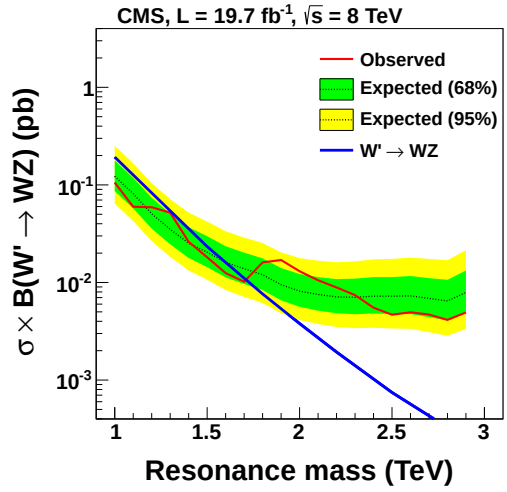
ATLAS has also released several conference notes using an incomplete 2015+2016 dataset, one for the fully hadronic channel [145] and two covering the semi-leptonic channels [12, 14]. The limits set by these single channel searches are generally the same or weaker than the 2015 combined analysis. The analysis performed for the fully hadronic note was updated with the full 2016 dataset and serves as the basis of the analysis presented here.

## 7.2 Software Framework

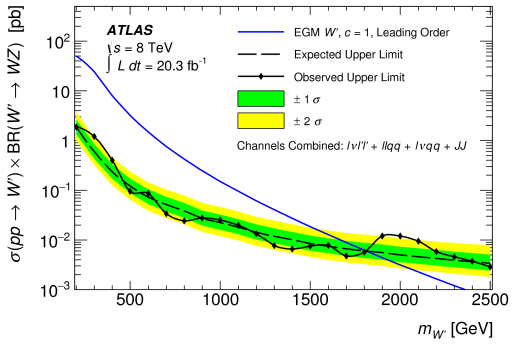
The primary software framework used for this analysis was the CxAOD (Calibrated xAOD) framework [146], which has been co-developed by several analysis groups within ATLAS. The CxAOD framework was used to perform data thinning, select events, and apply calibrations and systematic variations. Once CxAODs had been created, they were read using the PlotMaker histogramming tool.




(a)




(b)




(c)

Figure 7.2: The limits on diboson excesses placed by ATLAS and CMS using run-1 data. (a) shows the limits placed on $W' \to WZ$ by the fully-hadronic ATLAS search, excluding $1.3 < m_{W'} < 1.5$ GeV [17]. (b) shows the equivalent CMS search, which excluded $1.0 < m_{W'} < 1.7$ TeV [144], and (c) shows the ATLAS combined result, excluded $m_{W'} < 1.8$ TeV [54].

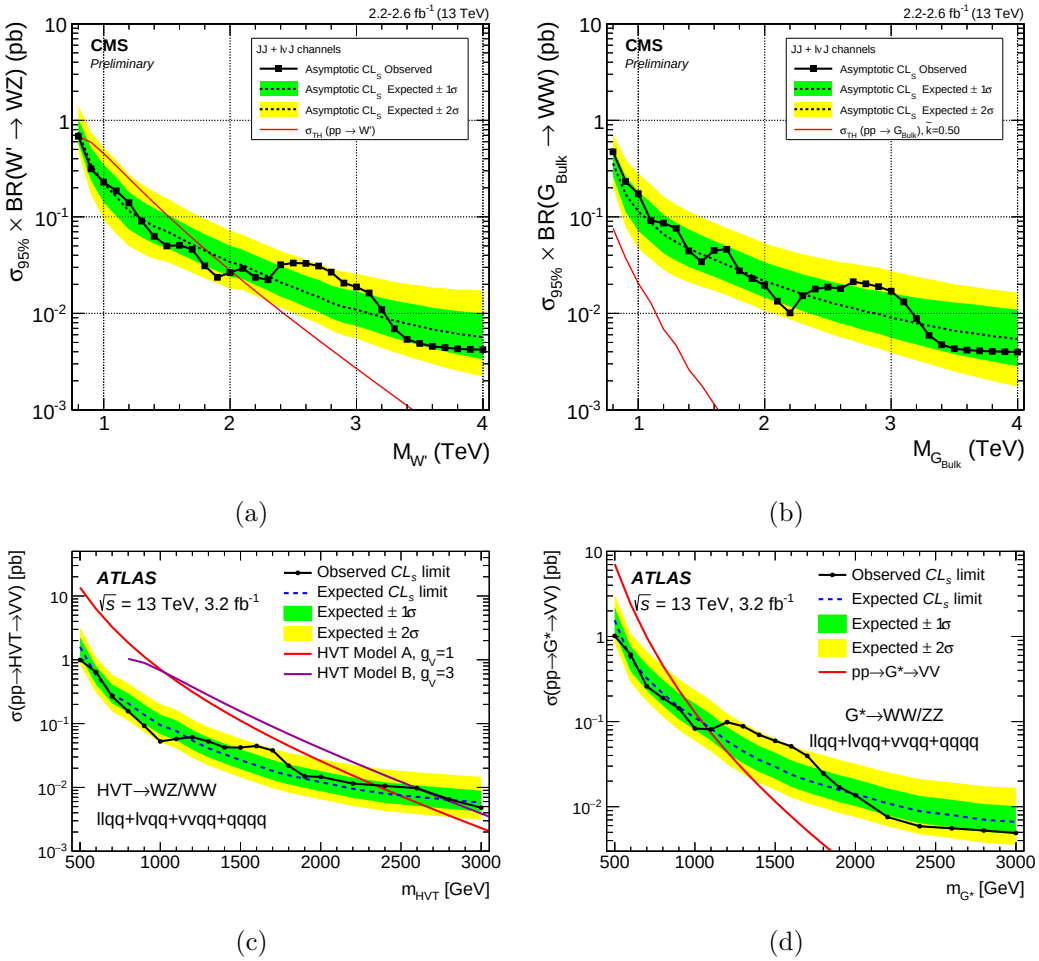



Figure 7.3: Limits on the HVT $W'$ (a,c) and RS graviton, $G_{RS}$, (b,d) by CMS and ATLAS using data recorded by the LHC in 2015 [13, 57]. The limits shown are not strictly comparable, the CMS search only used the $W' \to WZ$ decay mode to place limits on a HVT model B $W'$, whereas the ATLAS search placed limits on both model A and B using the full set of $W' \to VV$ decay modes. The graviton limits differ due to different choices in $\tilde{k}$ between the CMS and ATLAS analyses.

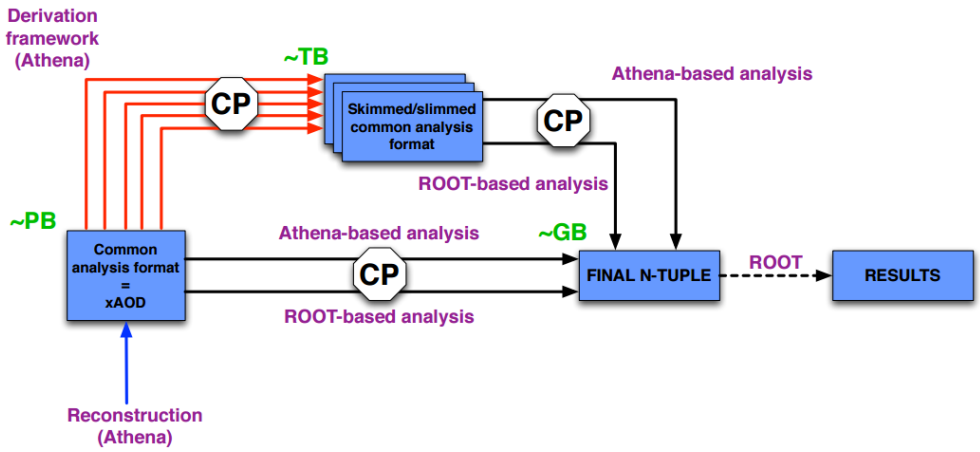



Figure 7.4: The ATLAS run-2 analysis model. The CxAOD framework sits between the common analysis format and the final-N-tuple (which in this context would be the CxAOD file format). Figure reproduced from [147].

### 7.2.1 CxAOD

The CxAOD framework provides a pipeline from ATLAS's standard reduced data format, DxAOD (Derived xAOD) [147], to a further reduced format, CxAOD as shown in figure 7.4. The framework applies calibration tools, generates systematic variations for MC samples, and can apply arbitrary event and object level selections. By sharing code between many analyses, the time needed to update and validate CP tools can be spread between analyses and the common framework means that common object definitions are used, making future combinations simpler. The framework was coded entirely in C++ and is built on top of EventLoop, a package for implementing generic processes involving an event loop, and RootCore, a compact build system which includes ROOT and many ATLAS tools.

Data reduction is achieved in three ways. First, events are only stored if they pass a preliminary (analysis-defined) selection. Secondly, even if an event passes, physics objects are only stored if they pass an object-level selection, these are shared between analyses and are typically very loose. For instance, large-R jets are stored only if they have $p_T > 200$ GeV and $|\eta| < 2.8$. Finally, most variables attached to each object are thrown away and only those relevant to each analysis are stored. A typical CxAOD event is $O(1kB)$ compared to $O(150kB)$ for the DxAODs. The full 2015+2016 dataset and associated MC datasets for the VVJJ analysis are around 800Gb in size.

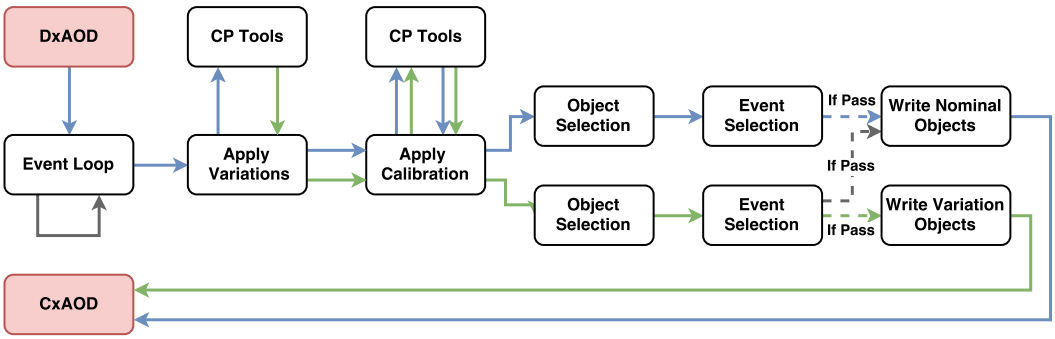



Figure 7.5: A simplified data flow model for the CxAOD framework, focusing on the event loop. The nominal (DxAOD) object data flow is shown in blue and the systematic variations in green. Conditional actions are indicated by a dotted line.

The core of the CxAOD framework consists of 5 packages:

**CxAODMaker:** The primary code for running CxAOD framework on DxAODs, including the event loop and code for applying CP tools and object level selections.
**CxAODTools:** Additional tools which can be used during both the reading and writing of CxAODs, e.g., event-level selection classes, pileup reweighting and overlap removal.
**CxAODReader:** The standard package for reading CxAODs. It was not used for the VVJJ analysis which had a dedicated reader package. See §7.2.2.
**FrameworkExe:** Contains the main executable and configuration settings. Replaced by FrameworkExe_VVJJ.
**FrameworkSub:** Utility code which does not fit elsewhere, including dataset definitions and bootstrap scripts.

and the analysis made use of three further derived or replacement packages:

**CxAODMaker_VHbb:** Object handlers derived from those in CxAODMaker with additional analysis-related functionality. This is shared between all diboson analyses and the exotic $VH \to X + bb$ searches (where $X$ is $ll, l\nu, qq$).
**CxAODTools_VHbb:** The VH and VVJJ analysis event selection classes.
**FrameworkExe_VVJJ:** Package for the VVJJ implementation executable.

### 7.2.2 PlotMaker

Once the CxAODs have been produced they must be read to produce the final analysis histograms. Unlike an n-tuple format, this is non-trivial as the files are stored in the xAOD format rather than a simple ROOT TTree. The "PlotMaker"

package produces histograms[2] and TTrees from input CxAOD or ROOT NTuple files.

The PlotMaker package is designed to be extremely flexible with several powerful capabilities:

- Output histograms and trees are defined in text files the output can be altered freely without recompiling
- Arbitrary event selections can be applied, including selections defined in the CxAOD framework.
- Per-histogram selections can be added and compiled at run-time.
- Standard ATLAS CP tools, such as pileup reweighting or taggers can be reapplied to the CxAOD.

The classes used by the PlotMaker package are:

**Looper** – Managers all other classes and runs the event loop.
**VarSet** – Creates and stores all variables and provides access to xAOD content. A ''variable'' in this context is a C++ std::function associated with a text string to enable run-time selection of variables.
**DataManager** – Handles samples and provides sample information.
**HistMaker** – Each instance handles a histogram definition file and automates creation/filling/writing of these histograms.
**TreeMaker** – Same as HistMaker but for TTrees.
**SelectionManager** – Manages per-histogram selections and provides them to any instances of HistMaker. Selections are mathematical expressions which are compiled from strings at run-time.
**EventSelector** – Handles all non-histogram selections and tells the Looper whether or not an event passed.
**WeightingTool** – Calculates weighting for each event including MC sample and event weights, and the ATLAS pileup reweighting tool.

The actual structure of the program is rather complex as it was designed to process events in a multithreaded fashion; it is shown in detail in figure 7.6.

[2]Which makes the title something of a misnomer.

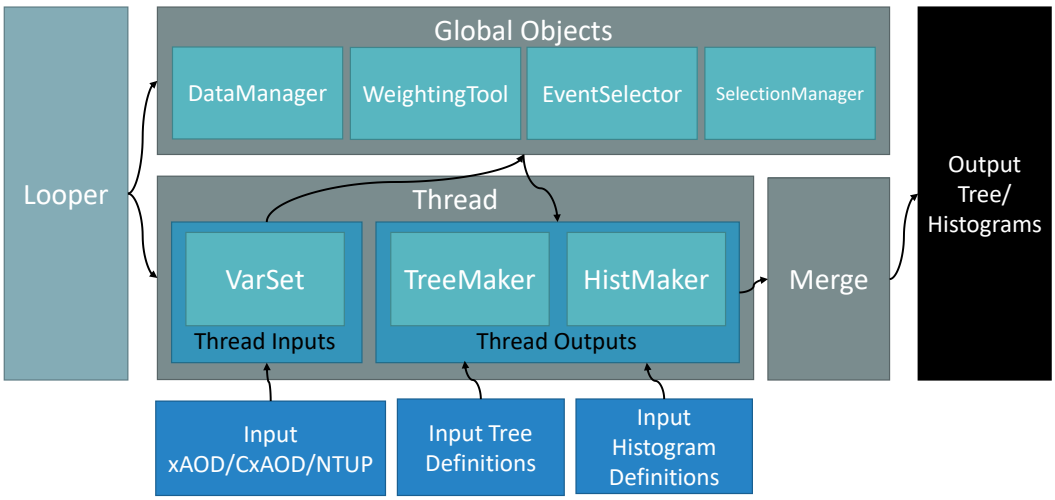



Figure 7.6: Data flow in the PlotMaker package. The work is divided into threads, each of which processes a set of events. When a thread is created it loads the input events and creates any output trees or histograms, it then iterates over the events and queries global classes as necessary. Once all threads have finished processing, their outputs are merged. Cross-class communication is managed in a functional fashion, where the state of global classes is immutable after initialisation.

## 7.3 Data and Monte-Carlo

### 7.3.1 Data Collection

The data used for this search was collected during the 2015 and 2016 $pp$ $\sqrt{s} = 13$ TeV runs of the LHC. Only runs with a 25ns bunch spacing were considered. For 2015 this corresponds to runs 276262-284484 and for 2016, runs 297730-311481. Good runs lists were applied to exclude data where the detector subsystems were not fully functional for physics collection. IBL-off runs were excluded for both 2015 and 2016, and toroid-off runs were excluded for 2015 only. The GRLs used that enforce these requirements were:

2015 Data: data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02_PHYS_StandardGRL_All_Good_25ns.xml

2016 Data: data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns_ignore_TOROID_STATUS.xml

The integrated luminosity corresponding to this dataset was estimated, using the methodology described in [118], as $3.2 fb^{-1}$ and $33.5 fb^{-1}$ for 2015 and 2016 data

respectively. Further to this selection, incomplete events, or those whose LAr or Tile quality is flagged as bad, were removed [119]. Finally, only data passing the lowest unprescaled large-R trigger in each run is considered, as will be described in §7.4.1.

### 7.3.2 MC Samples

While simulated events are not directly used for the background fit or statistical interpretation, the optimisation of the analysis strategy, estimation of sensitivity to specific models, and model limits were evaluated using simulated datasets. The events are simulated by a variety of generators using the processes described in §4.1. GEANT4 [115] is then used to simulate the interaction of the final state particles with the detector. Finally, the events are overlayed with additional minimum-bias interactions to simulate pileup within the detector. Once the simulation is complete the events are passed through the full ATLAS reconstruction software. Figure 7.7 compares the pileup distributions of the MC samples and collected data.

#### Background

#### QCD

The main expected background for this analysis is QCD multijet ''faking'' W and Z decays. The QCD samples are used during the optimisation of the analysis selection criteria but due to the very high rejection power of the boson tagging there are too few events passing to model the smoothly falling background, data-driven methods must be used instead. The samples were generated and showered with Pythia 8.186, with the A14 tune, in bins of the leading $p_T$ truth-jet (clustered with the anti-$k_T$ algorithm and $R = 0.6$). The sample bins are given in table 7.1. The PDF set used was NNPDF23LO [114]. The samples are weighted to produce a smoothly falling spectrum in $p_T$.

#### V+jets

Hadronically-decaying $W + jets$ and $Z + jets$ (collectively referred to as $V + jets$) samples were also generated for use in control region studies. $V + jets$ samples were produced with Herwig++ 2.7.1 [149], using the CTEQ6L1 PDF set [150], and were produced in bins of vector boson $p_T$, the samples produced are shown in table 7.2.

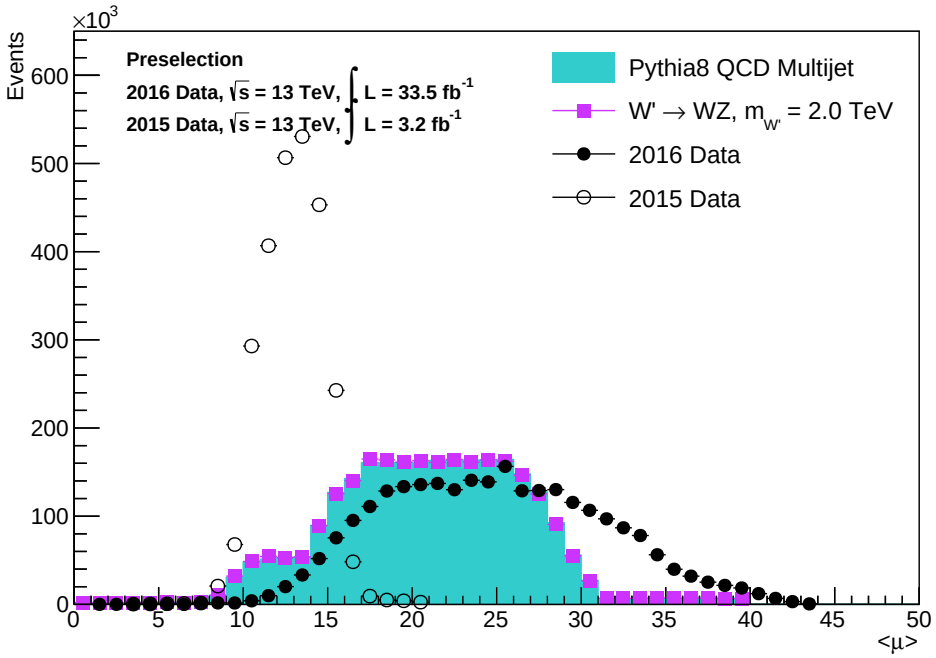



Figure 7.7: The average interactions per bunch crossing during the 2015 and 2016 runs compared to those simulated in the multijet and signal samples. There is a large difference between these datasets but the effect of pileup on the analysis selection was studied and found to have a negligible impact.

| | Truth jet $p_T$ [GeV] | | $\sigma$ | Filter | $n_{\text{evt}}$ | $L_{\text{int}}$ |
|---|---|---|---|---|---|---|
| Slice | Low | High | [fb] | Efficiency | [k] | $[fb^{-1}]$ |
| JZ3W | 160 | 400 | 2.6454e+10 | 3.1953e-4 | 1767 | 2.10e−1 |
| JZ4W | 400 | 800 | 2.5464e+8 | 5.3009e-4 | 1997 | 1.48e1 |
| JZ5W | 800 | 1300 | 4.5536e+6 | 9.2325e-4 | 1995 | 4.75e2 |
| JZ6W | 1300 | 1800 | 2.5752e+5 | 9.4016e-4 | 1997 | 8.25e3 |
| JZ7W | 1800 | 2500 | 16214 | 3.9282e-4 | 1990 | 3.12e5 |
| JZ8W | 2500 | 3200 | 625.05 | 1.0162e-2 | 2000 | 3.15e5 |
| JZ9W | 3200 | 3900 | 19.639 | 1.2054e-2 | 2000 | 8.45e6 |
| JZ10W | 3900 | 4600 | 1.196 | 5.8935e-3 | 2000 | 2.84e8 |
| JZ11W | 4600 | 5300 | 0.042258 | 2.7015e-3 | 1999 | 1.75e10 |
| JZ12W | 5300 | $\infty$ | 0.0010367 | 4.2502e-4 | 1808 | 4.10e12 |

Table 7.1: The cross-section and integrated luminosity of the QCD multijet samples [148]. Table taken from [143].

| Sample | Truth V $p_T$ [GeV] Low | High | $\sigma \times BR$ $[fb]$ | Filter Efficiency | $n_{\text{evt}}$ | $L_{\text{int}}$ $[fb^{-1}]$ |
|---|---|---|---|---|---|---|
| $W + jets$ | 280 | 500 | 13525.0 | 1.0 | 148000 | 10.94 |
| $W + jets$ | 500 | 700 | 920.71 | 1.0 | 29000 | 43.44 |
| $W + jets$ | 700 | 1000 | 175.3 | 1.0 | 29000 | 165.43 |
| $W + jets$ | 1000 | 1400 | 21.79 | 1.0 | 15000 | 1376.78 |
| $W + jets$ | 1400 | $\infty$ | 2.274 | 1.0 | 15000 | 8795.07 |
| $Z + jets$ | 280 | 500 | 5472.3 | 1.0 | 148000 | 13.71 |
| $Z + jets$ | 500 | 700 | 370.3 | 1.0 | 29000 | 54.01 |
| $Z + jets$ | 700 | 1000 | 70.598 | 1.0 | 29000 | 212.47 |
| $Z + jets$ | 1000 | 1400 | 8.8535 | 1.0 | 15000 | 1694.25 |
| $Z + jets$ | 1400 | $\infty$ | 8.8535 | 1.0 | 15000 | 1129.50 |

Table 7.2: The cross-section and integrated luminosity of the $V + jets$ sample slices. Taken from [143].

**Signal**

While the analysis is not targeted at a specific model, three benchmark models were used to estimate the signal efficiency and place concrete limits. These are the Heavy Vector Triplet (HVT) and bulk RS Graviton models, discussed in §3.4, and an additional generic heavy scalar model.

**Heavy Vector Triplet**

Model A HVT samples are used as a benchmark for this search. Model A corresponds to the addition of a new mass-degenerate multiplet, weakly coupled to the Standard Model fields, that decays to Standard Model fermions and the W, Z and Higgs boson. The branching ratios to $WW$, $WZ$, $WH$ and $ZH$ are all approximately 2%. Benchmark samples for 1.0 TeV $< m_{W'}, m_{Z'} <$ 5.0 TeV were generated with $g_V$, the coupling strength of the new multiplet, set to 1.0. With this assumption, the widths of the $W'$ and $Z'$ are around 2.5%. The matrix element calculations were performed using MadGraph 2.2.2 [151], integrated with Pythia 8.186, using NNPDF23LO PDF sets and A14 tune, for showering and hadronisation. Table 7.3 lists the $W'$ and $Z'$ samples generated.

**RS Graviton**

The analysis also sets limits on a spin-2 bulk RS graviton as described in §3.4.1. For the samples generated, the coupling constant $k/\overline{M}_{planck}$ was set equal to 1.0, giving a width of $\sim$ 6% and a branching fraction to $WW$ and $ZZ$ of between 16% and

| $m_{Z'}$ [TeV] | $\sigma \times BR$ [fb] | $n_{\text{evt}}$ [k] | $L_{\text{int}}$ $[fb^{-1}]$ |
|---|---|---|---|
| 1.3 | 32.28 | 29 | 8.98e2 |
| 1.4 | 22.78 | 25 | 1.10e3 |
| 1.5 | 16.36 | 30 | 1.83e3 |
| 1.6 | 11.92 | 30 | 2.52e3 |
| 1.7 | 8.79 | 20 | 2.28e3 |
| 1.8 | 6.56 | 30 | 4.57e3 |
| 1.9 | 4.94 | 30 | 6.07e3 |
| 2.0 | 3.76 | 20 | 5.32e3 |
| 2.2 | 2.22 | 20 | 9.01e3 |
| 2.4 | 1.35 | 15 | 1.11e4 |
| 2.6 | 0.831 | 30 | 3.61e4 |
| 2.8 | 0.521 | 30 | 5.76e4 |
| 3.0 | 0.332 | 29 | 8.73e4 |
| 3.5 | 0.111 | 30 | 2.70e5 |
| 4.0 | 0.037 | 5 | 1.35e5 |
| 4.5 | 0.014 | 29 | 2.07e6 |
| 5.0 | 0.0050 | 15 | 3.00e6 |

(a) $Z' \to WW \to qqqq$

| $m_{W'}$ [TeV] | $\sigma \times BR$ [fb] | $n_{\text{evt}}$ [k] | $L_{\text{int}}$ $[fb^{-1}]$ |
|---|---|---|---|
| 1.3 | 69.79 | 20 | 2.87e2 |
| 1.4 | 49.47 | 15 | 3.03e2 |
| 1.5 | 35.79 | 30 | 8.39e2 |
| 1.6 | 26.08 | 30 | 1.15e3 |
| 1.7 | 19.33 | 25 | 1.29e3 |
| 1.8 | 14.46 | 30 | 2.07e3 |
| 1.9 | 10.93 | 28 | 2.56e3 |
| 2.0 | 8.34 | 30 | 3.60e3 |
| 2.2 | 4.95 | 30 | 6.06e3 |
| 2.4 | 3.01 | 29 | 9.63e3 |
| 2.6 | 1.87 | 20 | 1.07e4 |
| 2.8 | 1.17 | 30 | 2.56e4 |
| 3.0 | 0.747 | 25 | 3.35e4 |
| 3.5 | 0.251 | 15 | 5.98e4 |
| 4.0 | 0.087 | 30 | 3.45e5 |
| 4.5 | 0.031 | 30 | 9.68e5 |
| 5.0 | 0.011 | 25 | 2.27e6 |

(b) $W' \to WZ \to qqqq$

Table 7.3: HVT $W'$ and $Z'$ simulation sample slices. Tables from [143].

18.5% depending on $m_{G_{RS}}$. The samples generated were for 1.0 TeV $< m_{G_{RS}} <$ 5.0 TeV and are described in detail in table 7.4. The graviton samples were produced using MadGraph 2.2.2 interfaced to Pythia 8.186 using the NNPDF23LO PDFs and A14 tune.

| $m_{G_{RS}}$ [TeV] | $\sigma \times BR$ [fb] | $n_{\text{evt}}$ [k] | $L_{\text{int}}$ [$fb^{-1}$] |
|---|---|---|---|
| 1.3 | 15.08 | 29 | 1.92e3 |
| 1.4 | 9.03 | 30 | 3.32e3 |
| 1.5 | 5.76 | 28 | 4.86e3 |
| 1.6 | 3.90 | 30 | 7.69e3 |
| 1.7 | 2.64 | 30 | 1.14e4 |
| 1.8 | 1.79 | 30 | 1.68e4 |
| 1.9 | 1.25 | 30 | 2.40e4 |
| 2.0 | 0.87 | 30 | 3.45e4 |
| 2.2 | 0.46 | 30 | 6.52e4 |
| 2.4 | 0.24 | 30 | 1.25e5 |
| 2.6 | 0.133 | 30 | 2.26e5 |
| 2.8 | 0.073 | 30 | 4.11e5 |
| 3.0 | 0.041 | 30 | 7.32e5 |
| 3.5 | 0.0107 | 29 | 2.71e6 |
| 4.0 | 0.0030 | 21 | 7.00e6 |
| 4.5 | 0.00082 | 30 | 3.66e7 |
| 5.0 | 0.00023 | 30 | 1.30e8 |

(a) $G_{RS} \to WW \to qqqq$

| $m_{G_{RS}}$ [TeV] | $\sigma \times BR$ [fb] | $n_{\text{evt}}$ [k] | $L_{\text{int}}$ [$fb^{-1}$] |
|---|---|---|---|
| 1.3 | 8.16 | 28 | 3.42e3 |
| 1.4 | 5.03 | 29 | 5.77e3 |
| 1.5 | 3.10 | 30 | 9.68e3 |
| 1.6 | 2.10 | 30 | 1.43e4 |
| 1.7 | 1.42 | 30 | 2.14e5 |
| 1.8 | 0.96 | 30 | 3.13e4 |
| 1.9 | 0.67 | 29 | 4.33e4 |
| 2.0 | 0.47 | 30 | 6.38e4 |
| 2.2 | 0.25 | 29 | 1.16e5 |
| 2.4 | 0.132 | 30 | 2.27e5 |
| 2.6 | 0.068 | 28 | 4.12e5 |
| 2.8 | 0.039 | 28 | 7.18e5 |
| 3.0 | 0.024 | 30 | 1.25e6 |
| 3.5 | 0.0057 | 28 | 4.91e6 |
| 4.0 | 0.0016 | 30 | 1.88e7 |
| 4.5 | 0.00044 | 30 | 6.82e7 |
| 5.0 | 0.00015 | 30 | 2.00e10 |

(b) $G_{RS} \to ZZ \to qqqq$

Table 7.4: $G_{RS}$ simulation sample slices. Tables from [143].

### 7.3.3 Heavy Scalar Model

Limits are also placed on a generic heavy spin-0 (scalar) boson produced via gluon-gluon fusion and subsequently decaying to either a WW or ZZ final state. $ggH \to ZZ \to qqqq$ and $ggH \to WW \to qqqq$ samples were produced using POWHEG [152–154] in a mass range of $1200 < m_H < 3000$ GeV. These were used for estimating the efficiency of the event selection for a scalar signal but were not used in the analysis optimisation. The analysis assumes that the decay width is negligible compared to the detector resolution, giving a mass resolution of 2.5% for WW and 3.5% for ZZ

states.

## 7.4 Preselection

A set of selections are applied to ensure that events considered by the analysis do not suffer from data quality defects and are on the trigger efficiency plateau. Additionally, events which pass the basic selection cuts for $VV \to llqq, l\nu qq, \nu\nu qq$ analyses are rejected to maintain orthogonality.

### 7.4.1 Trigger

Events are initially selected by the two-level ATLAS trigger system. For each year of data-taking, the lowest unprescaled large-R jet trigger was used. This corresponds to the triggers HLT_j420_a10_lcw_L1J100 in 2016 and HLT_j360_a10r_L1J100 in 2015. These are high level jet trigger chains with energy thresholds of 420 GeV and 360 GeV respectively. The 2016 trigger uses anti-$k_T$, R=1.0 jets (a10) reconstructed from calorimeter clusters with a local cluster weighting (lcw) calibration scheme applied, the 2015 trigger uses the same trigger algorithm with calibrated $R = 0.4$ anti-$k_T$ jets as inputs (a10r) and no further calibration is applied. Both triggers are seeded by a 100 GeV L1 jet trigger (LJ100).

The efficiency of the 2016 trigger was measured relative to a prescaled trigger using the same algorithm but with a 260 GeV energy threshold, HLT_j260_a10_lcw_L1J75. This lower threshold trigger can be regarded as 100% efficient for our signal in the region of interest (jets with $p_T > 350$ GeV). Due to the absence of such a trigger during 2015 running, the 2015 trigger efficiency was measured with respect to a 200 GeV small-R trigger, HLT_j200, whose efficiency plateau was checked in simulation to be well below the region of interest, as shown in figure 7.8.

The relative efficiency of our trigger for events passing the signal preselection is shown as a function of $p_T$ and $m_{JJ}$ ($m_{J_1J_2} = \sqrt{(E_1+E_2)^2 - (\vec{p}_{T,1}+\vec{p}_{T,2})^2}$) in figure 7.9. By applying a cut on the leading jet $p_T$ at 450 GeV and restricting the search range to $m_{JJ} > 1.1$TeV, the analysis (including sidebands) is restricted to the region where both triggers are $\approx 100\%$ efficient.

Figure 7.9 shows that there are a few bins above the jet threshold of 450 GeV to have minor inefficiencies ($< 1\%$) for values of $p_T$. These are caused by inefficiencies in the level 1 small-R jet trigger that seeds the chosen HLT chain. However, these

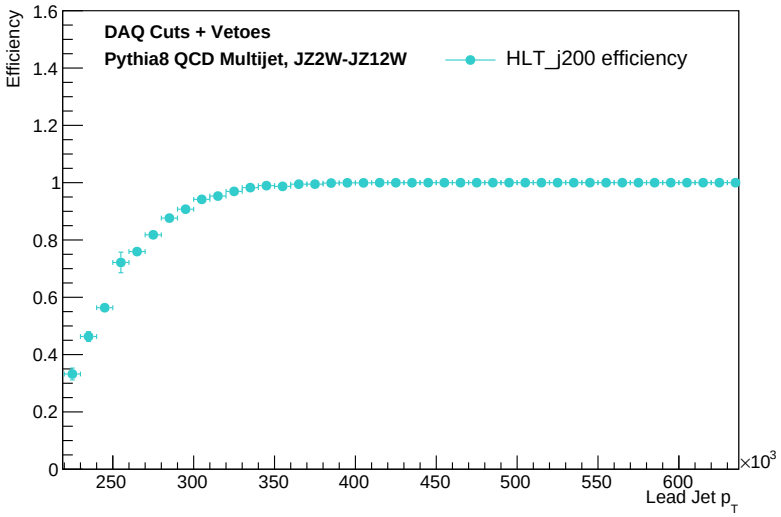



Figure 7.8: The efficiency of the HLT_j200 trigger, as a function of leading large-R jet $p_T$, measured in the QCD background samples. The trigger is fully efficient by 360 GeV and can be used for measuring the HLT_j360_a10r_L1J100 efficiency plateau.

inefficiencies are negligible and are removed altogether by applying boson tagging (see figure 7.10).

### 7.4.2 Derivation

To reduce the ATLAS dataset to a manageable size it is filtered through the ATLAS derivation framework [147] before the analysis takes place. The data is reduced through the rejection of events, removal of objects within the events and removal of attributes attached to objects. This is very similar in principle to the CxAOD data-reduction, but the cuts chosen are significantly looser to ensure that no relevant information is lost. When passing through the framework, an event is saved if it meets any of the following criteria:

- $VV$: >= 2 anti-$k_T$, $R = 1.0$ jets with $p_T > 100$ GeV, $|\eta| < 2.8$ and $m_J > 30$ GeV and passes a selected jet trigger.
- High $p_T$ $VV$: >= 2 anti-$k_T$, $R = 1.0$ jets with $p_T > 1000$ GeV and $|\eta| < 2.8$ and passes a selected jet trigger.
- $V + \gamma$: >= 1 anti-$k_T$, $R = 1.0$ jets with $p_T > 100$ GeV, $|\eta| < 2.8$ and $m_J > 30$ GeV, >= 1 photon or electron with $p_T > 100$ GeV and passes a selected photon

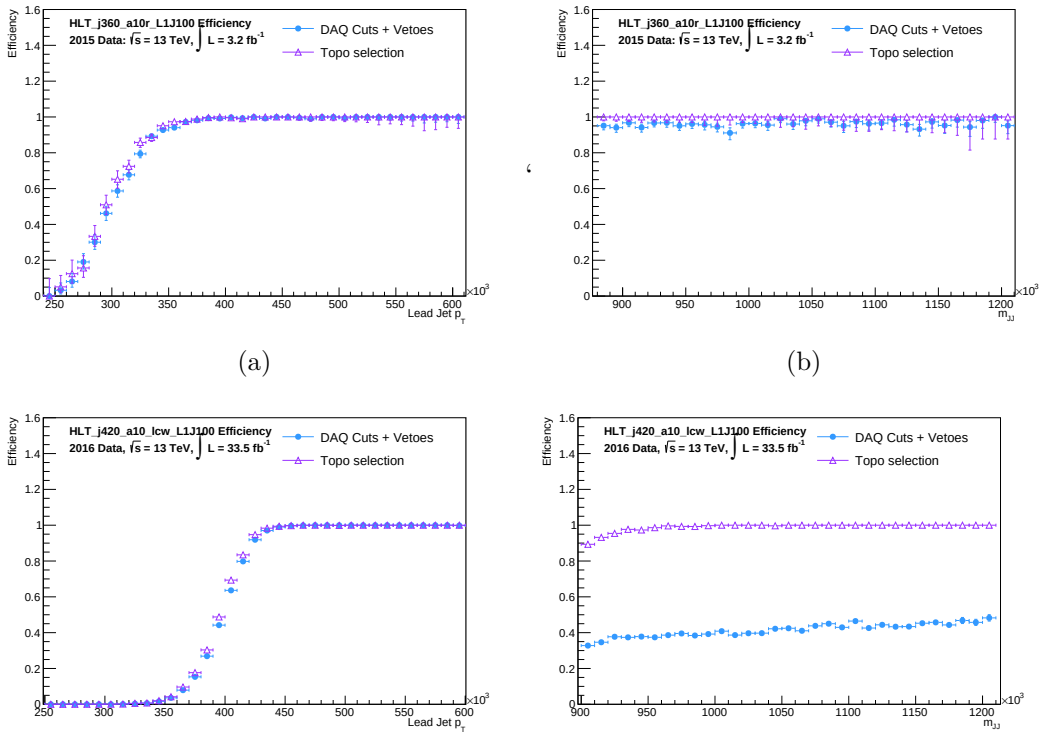



Figure 7.9: The efficiency of the HLT_j360_a10r_L1J100 (a,b) and HLT_j420_a10_lcw_L1J100 (c,d) triggers measured in data in 2015 and 2016 respectively. The trigger efficiency for data passing a partial preselection and the topological selection presented in §7.5.1 (each with the trigger, lead jet $p_T$, and $m_{JJ}$ cut removed) is shown as a function of lead jet $p_T$ and $m_{JJ}$. The 2015 trigger efficiency is measured with respect to data passing the HLT_j200 small-R jet trigger and the 2016 trigger efficiency is measured with respect to data passing the HLT_j260_a10_lcw_L1J75 trigger.

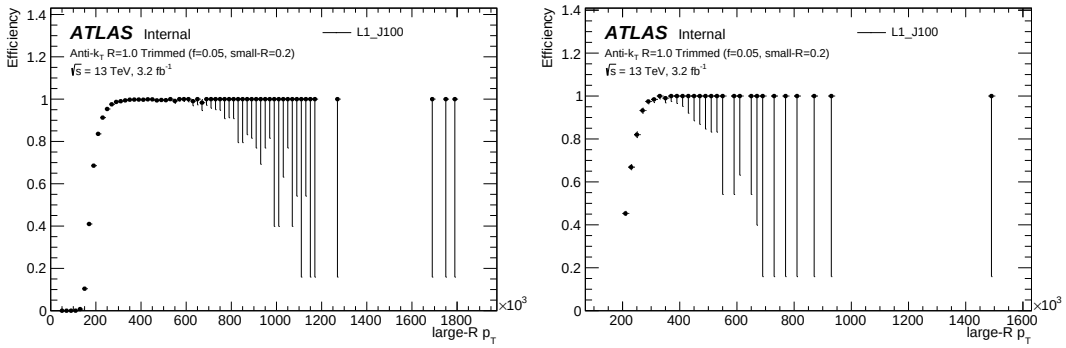
trigger[3].

The primary 2015/2016 jet triggers used by the derivation are:

- HLT_j360_a10r_L1J100
- HLT_j360_a10_lcw_sub_L1J100
- HLT_j400_a10_lcw_sub_L1J100
- HLT_j420_a10_lcw_sub_L1J100
- HLT_j460_a10_lcw_sub_L1J100

There were additional trigger chains included in the derivation but many were re-

[3]The derivation is shared with the $V + \gamma$ search, hence the additional allowance for a jet + photon event.




(a) Fail W/Z Tag

(b) Pass W/Z Tag

Figure 7.10: The efficiency of L1_J100 measured in 2015 data relative to an ''or'' of jet triggers including small-R and HT chains with lower $p_T$ level 1 seeds. The plots show the efficiency of this trigger with respect to $p_T$ for leading jets which fail (a) pass (b) a boson tag.

dundant, as they were not run during data collection, or targeted at $V + \gamma$ events. A large spread of allowed triggers was picked to ensure that the derivation would not lose data if the trigger menu were changed. The small-R jet and HT triggers are used for studies into alternative methods of triggering on large-R dijet events.

In addition to the event selection, individual tracks and calorimeter clusters are removed if they do not meet certain criteria. Tracks are kept if they have $p_T > 400$ MeV, $d_0 < 1.5$, $z_0 \sin\theta \leq 1.5$ and are matched to one of

- an electron with $p_T > 10$ GeV and $|\eta| < 2.8$,
- a photon with $p_T > 10$ GeV and $|\eta| < 2.8$,
- a muon with $p_T > 10$ GeV, and $|\eta| < 2.8$,
- a small-R anti-$k_T$ jet with $p_T > 15$ GeV and $|\eta| < 2.8$,
- a large-R anti-$k_T$ jet with $p_T > 150$ GeV and $|\eta| < 2.8$.

Calorimeter clusters are kept only if they are associated with an large-R anti-$k_T$ jet with $p_T > 150$ GeV and $|\eta| < 2.8$.

A further reduction is achieved by removing variables from objects where only a small amount of information is required (typically only the variables needed to calibrate them). This reduction is applied to the electron, muon, MET, track particle and vertex containers. Finally, any trigger data not associated with jet, electron or photon triggers is removed. The effect of the derivation on data and MC samples is shown in table 7.5.

| MC Sample | DSID | xAOD Events | DxAOD Events | Efficiency |
|---|---|---|---|---|
| JZ3W | 361023 | 7,884,500 | 3,564,938 | 45% |
| $W'(2TeV)$ | 301267 | 50,000 | 47,624 | 95% |
| Data | Period | xAOD Events | DxAOD Events | Efficiency |
| Run 280423 | G | 26,065,895 | 462,344 | 1.8% |

Table 7.5: The number of events passing the EXOT3 derivation for data and MC samples, and the reduction in file size. Table taken from [143].

### 7.4.3 Jet Cleaning

Events are not used in the analysis if they fail a jet cleaning cut designed to remove badly reconstructed jets. These can be caused by, for example, calorimeter noise, beam-induced background and cosmic rays. Events are rejected if any of the event's small-R anti-$k_T$ jets with $p_T > 20$ GeV are categorised as ''BadLoose'' [116], which identifies fake jets based on calorimeter cell pulse shape and the pattern of energy deposition within the detector. This selection is $> 99.5\%$ efficient for real jets.

### 7.4.4 Lepton and MET Vetoes

In order to preserve orthogonality with diboson searches in other channels (i.e. decays to $llqq$, $l\nu qq$ and $\nu\nu qq$), vetoes are made on electrons, muons and missing $E_T$ that would pass the selection in their respective channels. ''Veto'' electrons and muons are required to have $p_T > 25$ GeV, $|\eta| < 2.5$ (2.47 for muons), and pass their respective ''medium'' identification and track isolation requirements [155, 156].

Veto electrons (muons) are also required to pass cuts on the transverse impact parameter, $|d_0| < 5.0$ (3.0) mm, the longitudinal impact parameter $|z_0 \sin\theta| < 0.5$ (0.5) mm and on the significance of the transverse impact parameter, $|d_0|/\sigma_{d_0} < 5$ (5).

Any event with one or more leptons passing these cuts is vetoed, as in any event containing $E_T^{miss} > 250$ GeV.

## 7.5 Event Selection

### 7.5.1 Topological Selection

Once the data has passed the described trigger and data quality selection, cuts are applied on the event topology to enhance the sensitivity of the search. The event selection cuts were optimised using the $W' \to WZ$ signal samples.

#### Jet Selection

A stronger version of the jet selection applied during the derivation is used. In addition to the trigger efficiency cuts, we require that the two leading trimmed $R = 1.0$ anti-$k_T$ jets have $p_T > 400$ GeV, $|\eta| < 2.0$ and $m_J > 50$ GeV after the energy and combined mass calibrations are applied. The trimming is applied as in §4.3.1 with $R_{sub} = 0.2$ and $f_{cut} = 0.05$ .

The $\eta$ cut ensures that the jets are contained within the tracking volume of the detector so that the tracking information so can be used for boson tagging and the mass calibration. The restriction to $m_J > 50$ GeV keeps the jets in the region where the mass calibration is well-defined.

#### $\Delta$Y

A cut is made on the rapidity difference between the two leading jets. For both $W'$ and graviton signals the main production mode is expected to be s-channel, giving a $\Delta Y$ distribution centered around 0. For the QCD background the primary production mode is t-channel exchange, which gives a $\Delta Y$ distribution peaked at a higher value, as shown in figure 7.11. The significance gained by this cut is shown in figure 7.12. Note that the cut is on *rapidity* rather than *pseudorapidity*, this is for historical reasons and in practice the cut is equivalent to $|\Delta\eta| < 1.2$ for jets able to pass the analysis selection, which will have $p_T > 450$ GeV and $m_J < 110$ GeV, corresponding to $p_T/m > 4.09$. The exact relationship between $\Delta Y$ and $\Delta\eta$ is shown in figure 7.13.

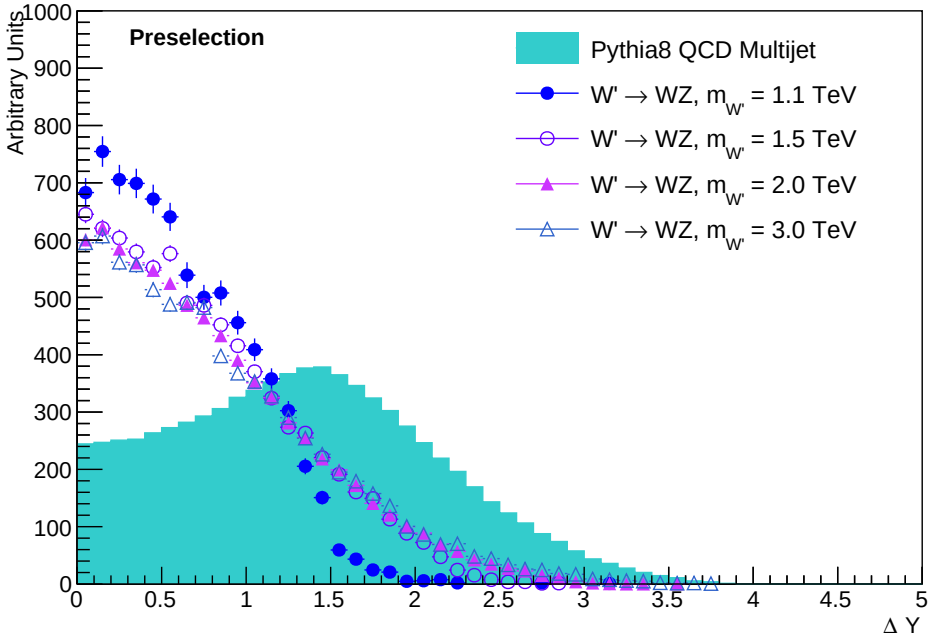


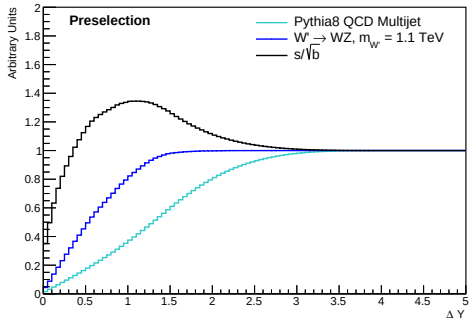

Figure 7.11: The $\Delta Y$ distribution for QCD multijet and $W'$ signal samples used for optimisation.

### $p_T$ Asymmetry

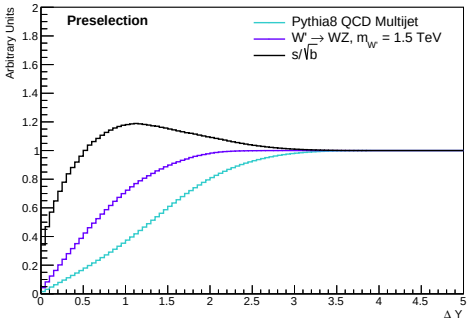
A heavy new particle should be produced at close to rest in the transverse direction, therefore the decay products, i.e. the two leading jets in the event, must be roughly balanced in $p_T$. To account for this, a cut is placed on the $p_T$ asymmetry of the leading jets defined as,

$$A_{p_T} = \frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}} < 0.15 \tag{7.1}$$

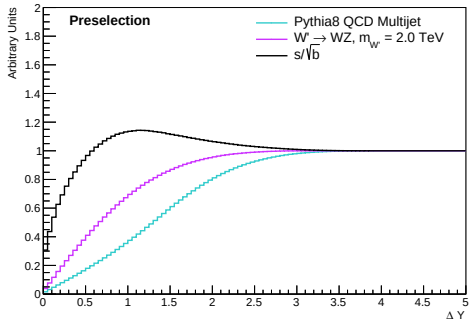
$A_{p_T}$ should be small for any well measured dijet event including both $W' \to WZ$ decay and QCD dijets, hence the main purpose of this cut is not to suppress the dijet background but to remove events with a mis-modelled or mis-measured jet. Figures 7.14 and 7.15 show that, as expected, the signal and background distributions are similar and any discriminating power provided by this cut is removed by the previous cut on $\Delta Y$.



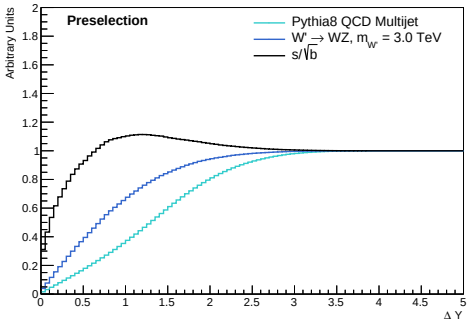

(a) $m_{W'} = 1.1$ TeV




(b) $m_{W'} = 1.5$ TeV




(c) $m_{W'} = 2.0$ TeV




(d) $m_{W'} = 3.0$ TeV

Figure 7.12: Plots show the proportion of background and signal samples passing preselection that also pass a cut on $\Delta Y$ corresponding to the x-axis value. Also shown is the ratio of passing signal events to the square root of passing background events. A cut of 1.2 was chosen based on the signal to background ratio and signal efficiency.

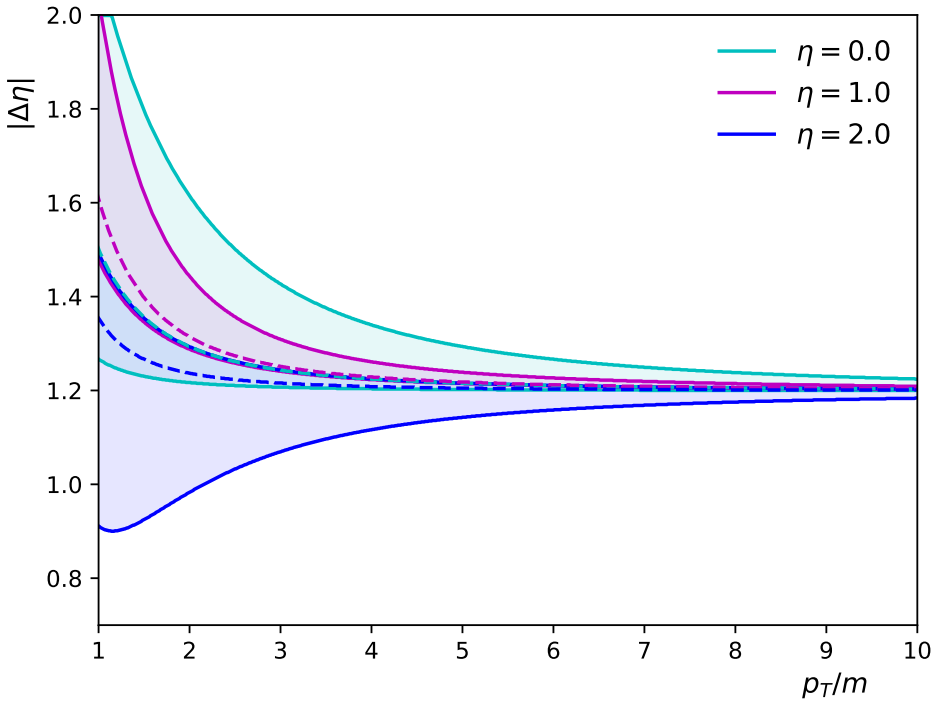
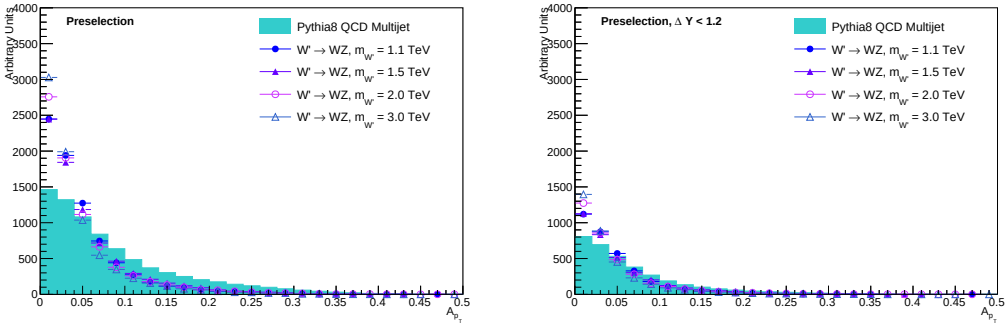
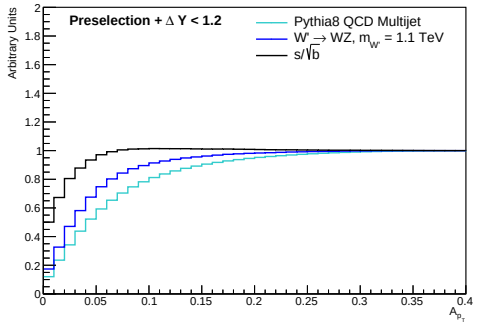



Figure 7.13: The cut $\Delta Y < 1.2$ as an equivalent cut on $\Delta\eta$ as a function of $p_T/m$ of the jet with highest $\eta$. Each shaded region represents the approximate range of possible values of $|\Delta\eta|$ as $p_{T_2}/m_2$ is varied between 0.4-2.4 $\times p_{T_1}/m_1$ (roughly the allowed range if both jets pass the boson mass and $p_T$ asymmetry cuts). The dotted lines show the value when $p_T/m$ of both jets is equal.

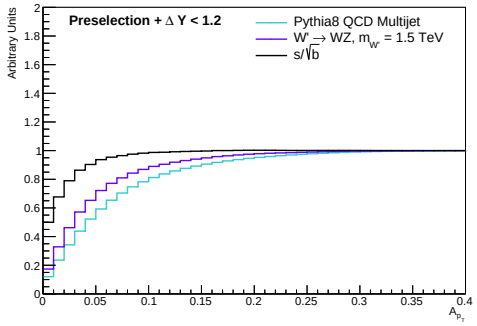



(a) (b)

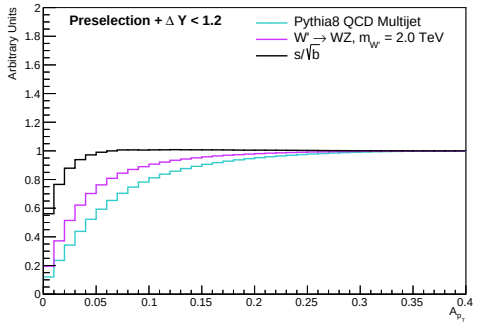
Figure 7.14: The $p_T$ asymmetry distribution for the QCD multijet and $W'$ signal samples before (a) and after (b) the $\Delta Y$ cut is applied. After the $\Delta Y$ cut the discrimination power of the cut is minimal.



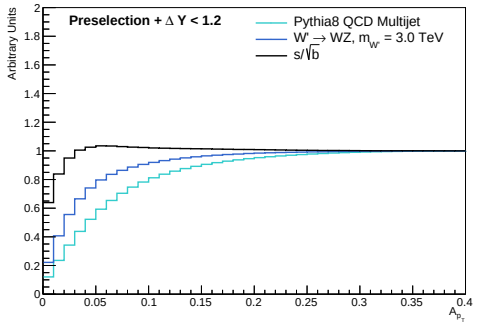

(a) $m_{W'} = 1.1$ TeV




(b) $m_{W'} = 1.5$ TeV




(c) $m_{W'} = 2.0$ TeV




(d) $m_{W'} = 3.0$ TeV

Figure 7.15: The percentage of background and signal samples passing the preselection and $\Delta Y$ cuts which also pass a cut on $A_{p_T}$ corresponding to the x-axis value. Also shown is the ratio of the fraction of passing signal events to the square root of the fraction of passing background events.

### 7.5.2 Boson Tagging

The jet substructure information explored in §4.3 can be used to provide discriminatory power between QCD and boson-initiated jets. These techniques have been used by both ATLAS [89, 92, 157] and CMS [158] and this analysis builds on the recommendations in [92]. After trimming, a jet can be tagged as a W or Z jet based on jet mass, substructure, and track multiplicity requirements. The mass and substructure tagging parameters were optimsed simultaneously for maximum QCD jet rejection at 50% signal acceptance using $W' \to WZ$ signal samples. The $n_{trk}$ cut was then optimised on the same samples using the $s/\sqrt{b}$ plots shown in figure 7.18.

The W and Z tag requirements are overlapping and a jet may be tagged as both a W and a Z jet. During the limit setting procedure the single signal channels (WW,WZ,ZZ) are treated as independent and populated based on these tags so, for instance, an event whose jets pass both the W and Z tag would appear in each channel. The combined channels, WW+ZZ and WW+WZ, are also considered as separate selections requiring, e.g., 2 W tags and/or 2 Z tags, and are not built by combining the single channel results. For the signal and mass sideband selections defined in §7.6 the two leading jets are sorted by mass. Jets sorted in this fashion are here-on referred to as the ‘‘first’’ and ‘‘second’’ jet rather than ‘‘leading’’ and ‘‘subleading’’.

#### Jet Mass

After trimming, the mass of a boson-initiated jet is likely to be close to that of the boson peak, especially when the combined mass is used, whereas a QCD jet will typically have a much lower mass. A jet passes the mass selection if its combined mass is within a $p_T$ dependent window in the region of the boson mass. The upper and lower mass boundaries, $m_{cut}^{low}(p_T)$, $m_{cut}^{high}(p_T)$ are defined by a 4D function fit to $W'$ MC samples. This fit uses the physically-motivated function

$$\sqrt{\frac{a_0^2}{(p_T - a_1)^2} + a_2^2(p_T - a_3)^2} \tag{7.2}$$

Here the first term is a first-order approximation of the uncertainty in mass due to the jet energy resolution, which is dominant for low $p_T$ jet and scales as $\sim 1/p_T$, and the second term characterises the uncertainty in angular resolution at high $p_T$.

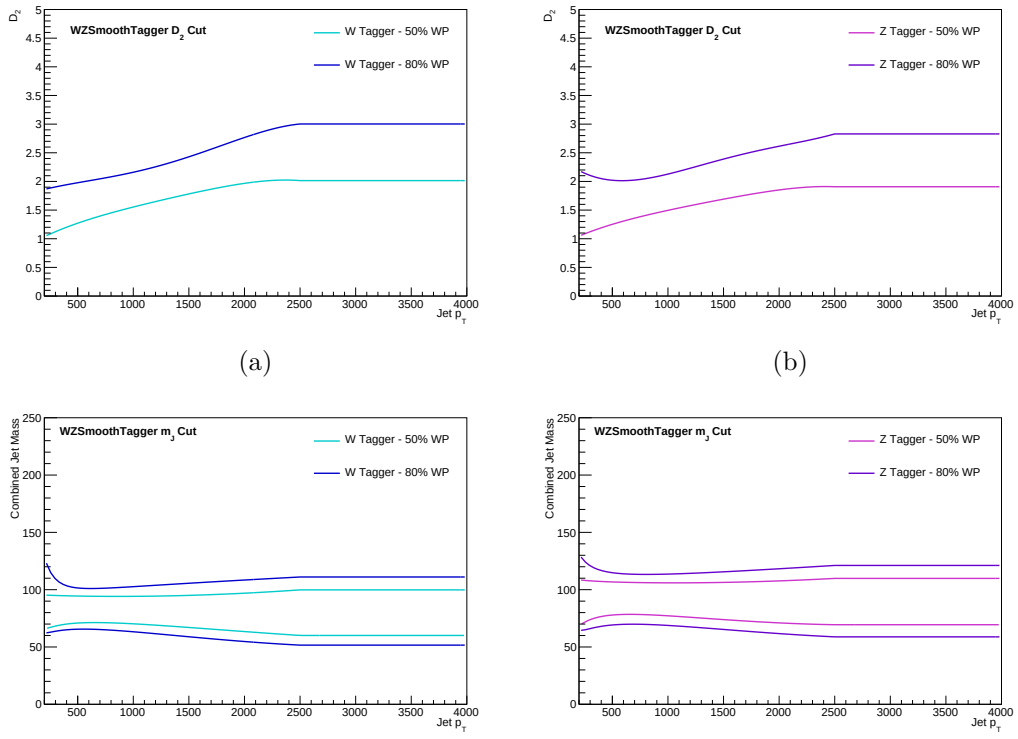


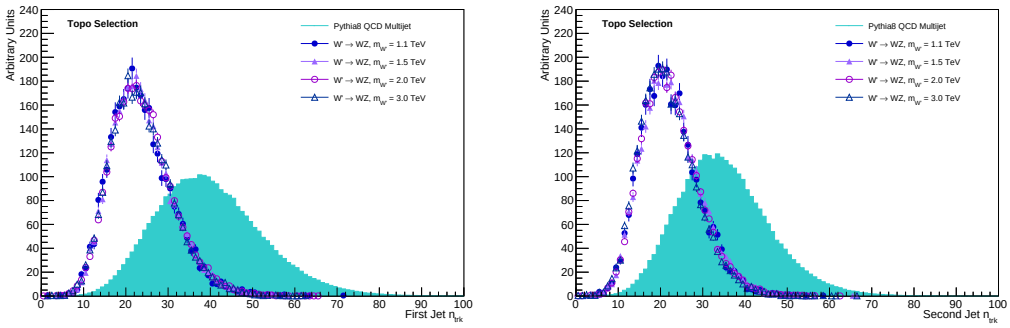

Figure 7.16: Boson tagger cut functions for the 50% and 80% signal acceptance working points. The mass and $D_2$ cuts were optimised simultaneously to achieve the maximum background rejection at each fixed working point. Plots (a) and (b) show the single-sided $D_2$ cut for W and Z jets, plots (c) and (d) show the double sided cut on combined mass for W and Z jets. Each cut function is replaced with a constant value for jets with $p_T > 2.5$ TeV.

### Jet Substructure

Since a boson jet will have a characteristic two-prong structure we use the $D_2^{\beta=1}$ variable, motivated in §4.3.2, to reject one-pronged background. Jets are rejected if they have $D_2 > f_{cut}(p_T)$ where $f_{cut}(p_T)$ is a $p_T$-dependent, 4D polynomial, fit to $W'$ MC samples. The fitted mass and substructure cut functions are shown in figure 7.16.

### Track Multiplicity

We use the number of tracks ghost-associated [159] with the ungroomed jet to further discriminate between bosons and other jets. The tracks are required to have $p_T >$




(a) (b)

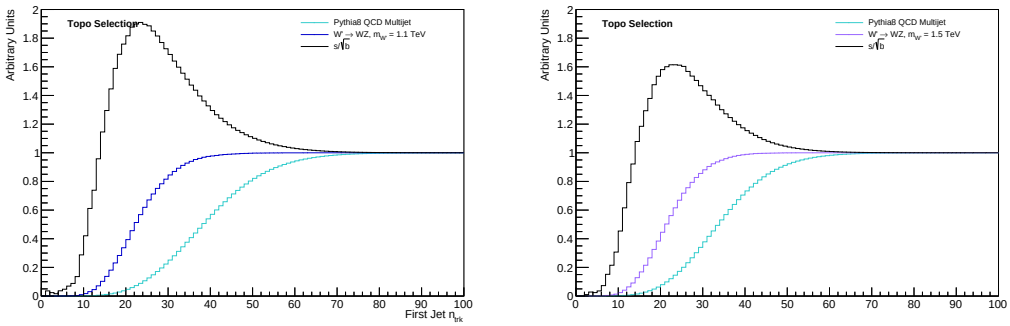
Figure 7.17: The $n_{trk}$ distribution for the first (a) and second (b) jet in QCD multijet and WZ signal samples after the topological selection has been applied.

500 MeV, $|\eta| < 2.5$, and be matched to the primary vertex. In addition to the colour factor effects described in §4.3.2, the emission rate of a jet increases with energy scale. For W/Z jets this scale is set by the mass of the boson, significantly lower than the QCD jet scale, which is set by the energy of the hard interaction that produced it. The $n_{trk}$ distributions for both QCD jets and W/Z are shown in figure 7.17 and the cut optimisation plots in figure 7.18 were used to set $n_{trk} < 30$ to maximise $S/\sqrt{B}$ for a variety of signal masses. The uncertainties on this cut were measured directly in data as will be described in §7.6.4

The first two steps of the W and Z taggers, the mass and substructure, correspond to the ATLAS recommended boson tagger and were optimised simultaneously to obtain constant 50% signal efficiency, independent of the jet $p_T$. The taggers reject approximately 98.5% of background jets, but this number varies with $p_T$. Since the tagger is applied twice, we accept 25% of simulated signal events while rejecting 99.98% of background events. The full topological selection and the jet mass and substructure cuts have a combined signal acceptance of $\sim$ 18% and background rejection of $\sim$ 99.998%, depending on the signal type and mass. After applying the track multiplicity criteria to both jets these become $\sim$ 12% and $\sim$ 99.9995% respectively.

### Pileup Dependence

The boson tagging recommendations were developed based on signal samples with a pileup distribution that differs from that of the data collected (as shown in figure 7.7). The impact of pileup on the mass and $D_2$ was found to be extremely small but




(a) (b)

Figure 7.18: $S/\sqrt{B}$ plots for the $n_{trk}$ cuts for the first (a) and second (b) jet. Note that since this is from a $W' \to WZ$ decay the first (second) jet signal sample contains is a nearly pure sample of Z(W)s. The $S/\sqrt{B}$ is maximised for values lower than the chosen 30 track cut but a higher value was chosen to account for the increase in $n_{trk}$ with pileup.

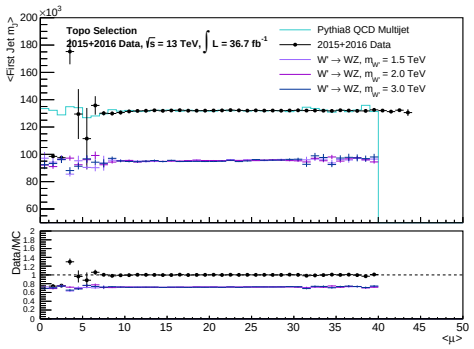
increasing pileup was seen to increase each jet's $n_{trk}$. Figure 7.19 shows the impact of pileup on the boson tagging variables in data and simulation.

#### Boson Polarisation

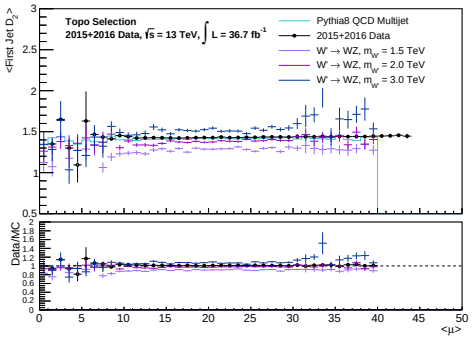
The boson tagging procedure and topology cuts are sensitive to the angular distribution of energy in each jet and the angular correlations between the jets respectively, as a result, the signal acceptance can vary strongly depending on the production mode and boson polarisation. The signal selection efficiency for a gluon (quark)-produced 2 TeV graviton was 12.3% (5.4%) if it decayed to longitudinally polarised Ws and 1.8% (5.2%) if it decayed to transversely polarised Ws. The large difference observed for the gluon-produced gravitons is primarily due to the 2nd jet falling outside the of the $\Delta Y$ cut. On a per-jet level, there is an approximately 20% difference in acceptance depending on the boson polarisation.

### 7.5.3 Selection Efficiencies

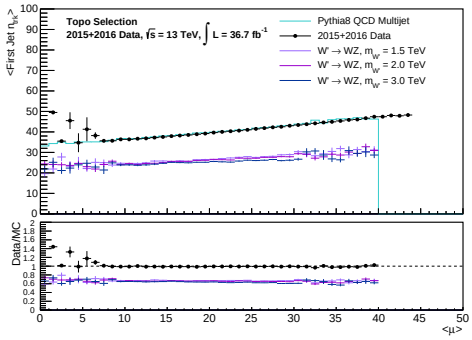
The dijet mass distribution for the QCD multijet samples is shown for each step of the event selection in figure 7.20. The selection efficiency of each cut on the background and signal datasets are shown in figure 7.21.




(a) First Jet $m_J$




(b) First Jet $D_2$




(c) First Jet $n_{trk}$

Figure 7.19: The mean value of $m_J$ (a), $D_2$ (b) and $n_{trk}$ (c) as a function of $\langle\mu\rangle$ for 2015+2016 data, QCD multijet and W' samples after the topological selection is applied. Only $n_{trk}$ exhibits any dependence on the pileup. The behaviour of these variables in the QCD multijet samples is entirely consistent with the behaviour observed in data.

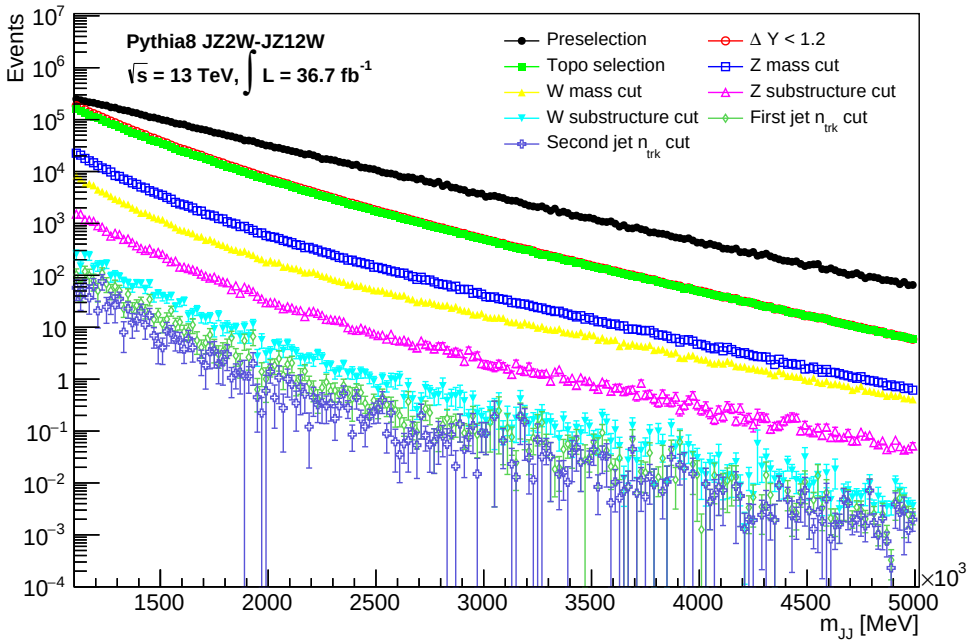
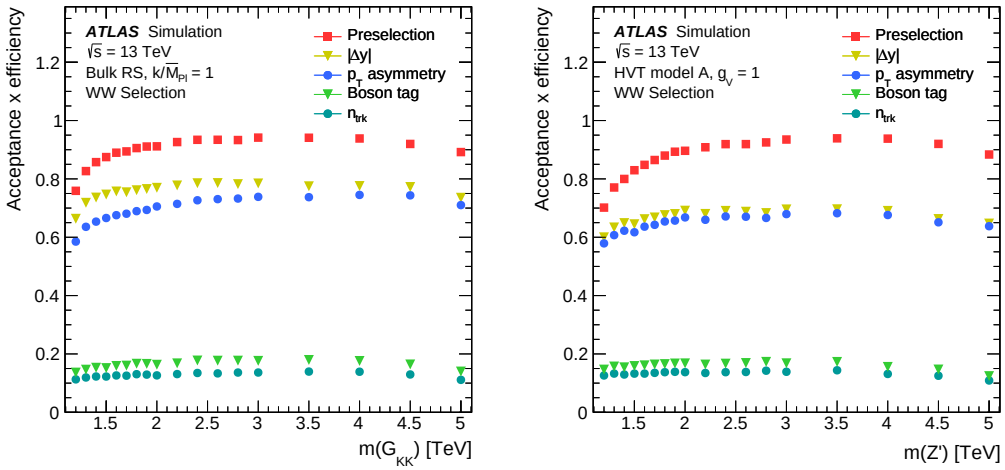


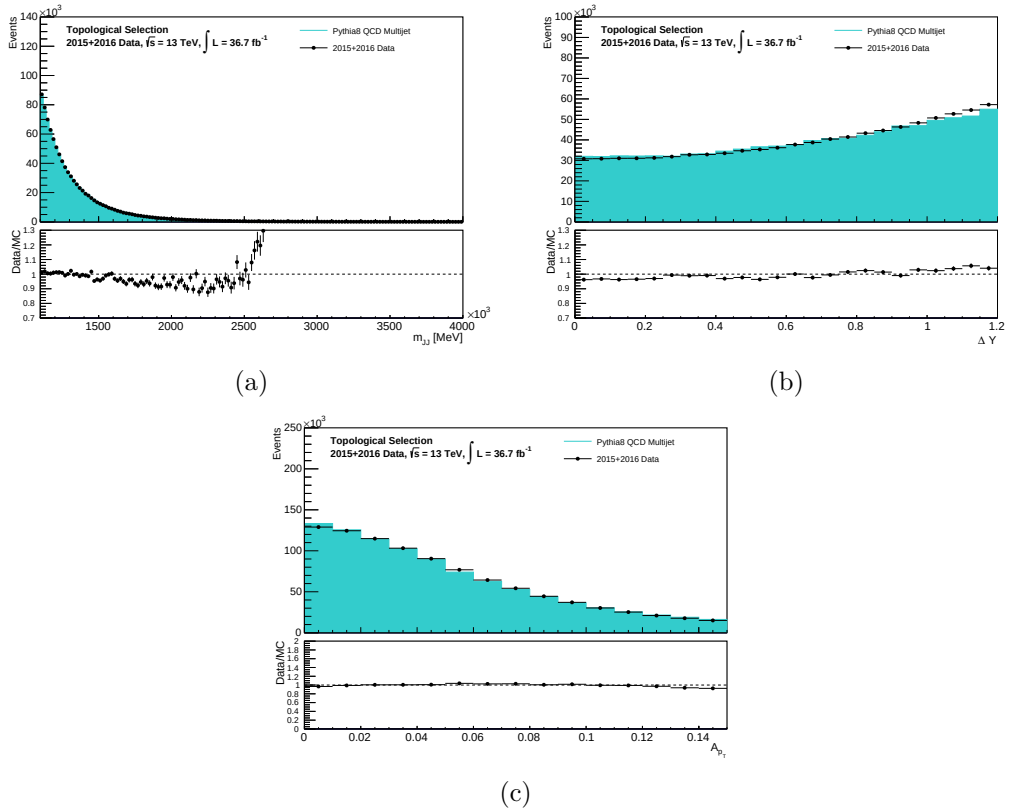

Figure 7.20: The dijet mass distribution for the QCD multijet samples at each step of the WZ selection after trigger, data quality and vetoes have been applied. No significant sculpting of the background occurs with any individual cut. The luminosity is scaled to the $36.7 fb^{-1}$ of collected data.




(a) Bulk RS Graviton Efficiencies

(b) HVT model A Efficiencies

Figure 7.21: The efficiency of each step in the selection relative to the derivation preselection for the RS Graviton (a) and HVT Model A (b) samples. Plots taken from [11].




(a) (b)

(c)

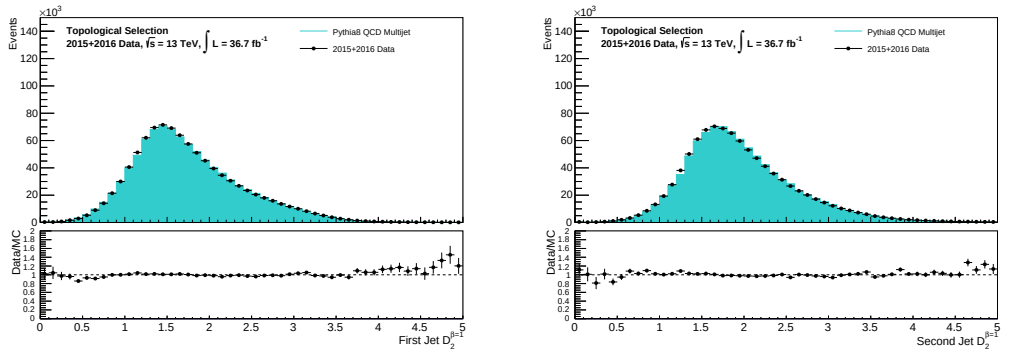
Figure 7.22: The dijet-mass (a), $\Delta Y$ (b), and $A_{pT}$ (c) distributions observed in the 2015 and 2016 data compared to those in QCD multijet simulation after the topological selection is applied. The Monte-Carlo distributions are scaled to match the number of observed data events.

## 7.6 Control Regions

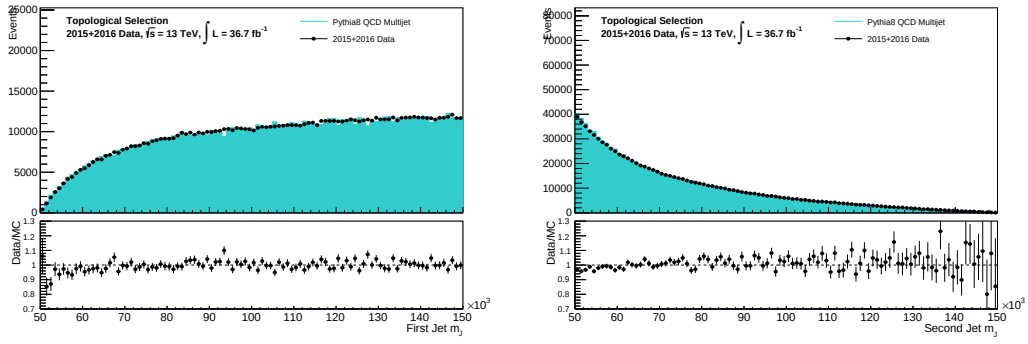
A set of control regions are used to validate the analysis strategy and MC simulation, in particular, the event selection, boson tagging and background fitting procedure, and estimate uncertainties. The regions used are presented in the following section.

### 7.6.1 Untagged

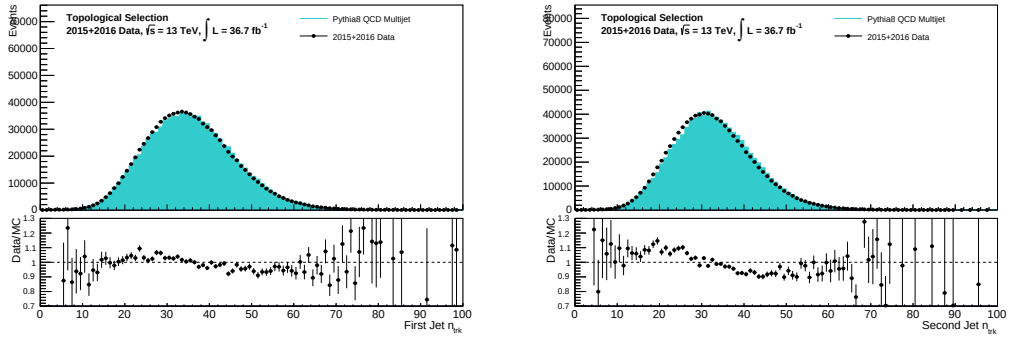
An untagged region, where only the topological selections are applied, is used to validate the MC modelling of the background before tagging is applied. The measured distributions are shown in figures 7.22 and 7.23.




(a)




(b)




(c)

Figure 7.23: The jet mass (a), $D_2$ (b), and $n_{trk}$ (c) distributions observed in the 2015 and 2016 data compared to those in QCD multijet simulation after the topological selection is applied. The monto-carlo distributions are scaled to match the number of observed data events. Good agreement is seen between all distributions, with the exception of $n_{trk}$.

### 7.6.2 Mass Sidebands

Mass sidebands with a modified boson tag selection are used to validate the background fit in §7.7 without unblinding the signal region. To get as close as possible to the signal region, the sideband selection applies the topological selection, $D_2$ and $n_{trk}$ cuts. To maximise the statistics, the $D_2$ cut is set at a looser, 80% signal efficiency, working point. Jets are categorised by exclusive low-mass or high mass tags, defined as

- **Low Mass:** 50 GeV $< m_J < m_{cut}^{low}(p_T)$, pass the 80% $D_2$ W-jet working point, and $n_{trk} < 30$
- **High Mass:** $m_{cut}^{high}(p_T) < m_J <$ 140 GeV, pass the 80% $D_2$ Z-jet working point, $n_{trk} < 30$, and no $R = 0.2$ sub-jets tagged as b-jets.

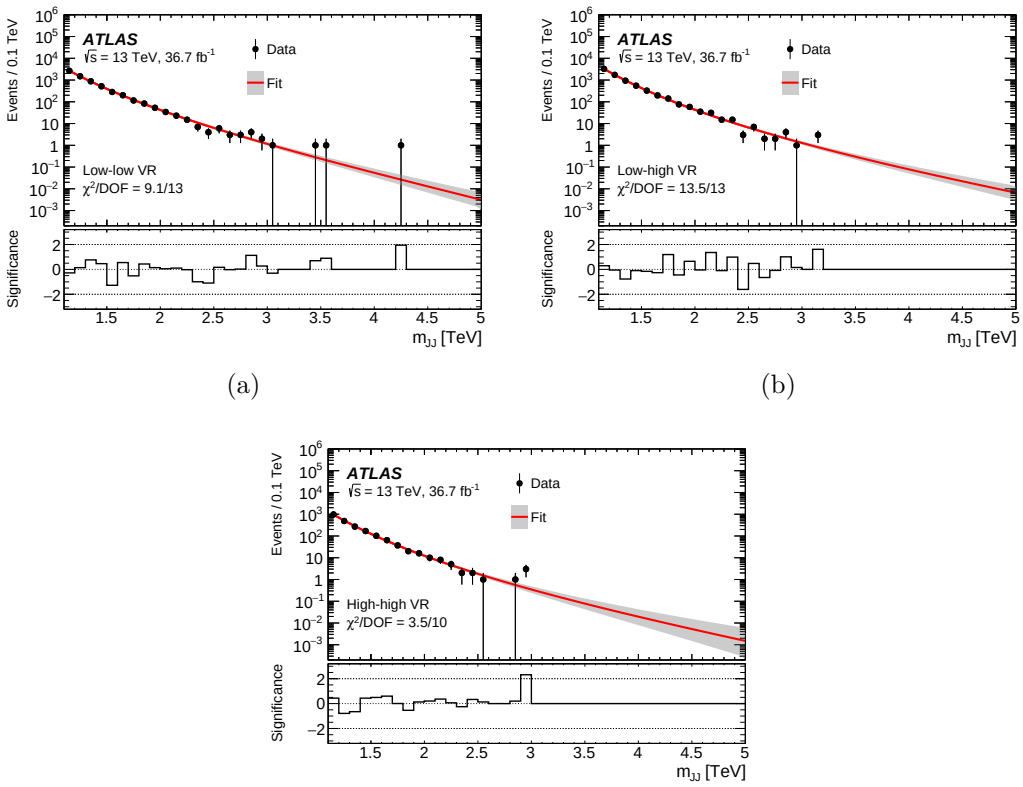
The b-jet veto applied to high-mass jets is to avoid the possibility of unblinding the signal region of $X \to VH \to qqbb$ analysis. As with the main selection, the two leading jets are sorted by mass. Three mass sidebands are formed from these definitions, LowLow, LowHigh, and HighHigh. The dijet mass distribution for each is shown in figure 7.24.

### 7.6.3 Partial Tag

Control regions are defined where the boson tag procedure is only partially applied. Two new signal regions are defined, the $D_2^{loose}$ region where an 80% efficiency working point is chosen, and the $n_{trk}^{loose}$ region where no $n_{trk}$ cut is applied.

### 7.6.4 V + Jets

A V+jets control region is used to estimate the uncertainty on the number of tracks ghost-associated with boson and quark-initiated jets. The basic procedure is as follows: first fit a signal rate model, as a function of $n_{trk}$ and mass, to a truth-tagged sample of V+jets generated using Herwig as described in 7.3.2. A signal+background fit is then made to a V+jets enriched data sample in 8 bins of $n_{trk}$, with 42 free parameters, $8 \times 5$ parameters for the background fit, an overall V+jets rate parameter and a Data/MC $n_{trk}$ scaling parameter. The closure of this procedure was estimated in MC and found to give an $n_{trk}$ scaling of $1.00 \pm 0.03$. The uncertainty here is estimated by varying the mass peaks and ratio of $W/Z$ in the signal fit.



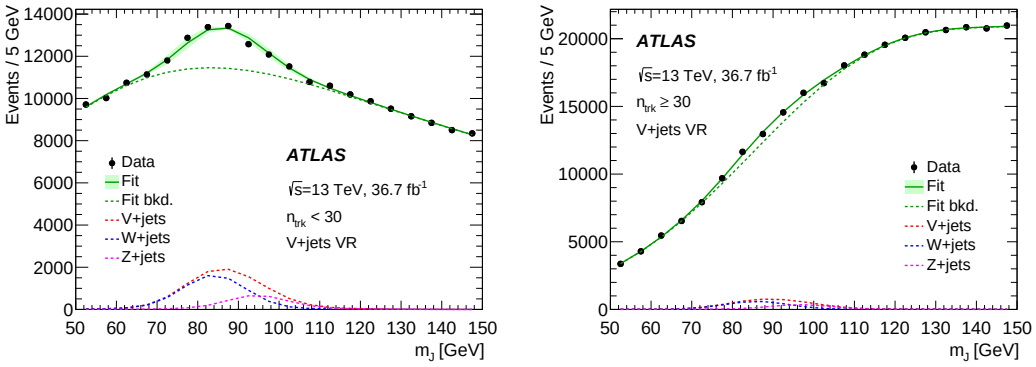

(a) (b)

(c)

Figure 7.24: The observed dijet mass distribution in the HighHigh, HighLow and LowLow mass sidebands. The distributions show that the fit function performs well in data. Reproduced from [11].

The V+jets enriched sample is obtained by requiring at least two large-R jets with $|\eta| < 2.0$. The leading jet is required to have $p_T > 600$ GeV and pass the W or Z $D_2$ cut. The fit is restricted to the region $50 < m_J < 150$ GeV and $n_{trk} < 40$.

This procedure measured a scaling in the number of tracks between MC and data of $1.03 \pm 0.03(stat) \oplus 0.03(sys) \oplus 0.03(closure) = 1.03 \pm 0.05$ which corresponds to a scaling in the efficiency of the $n_{trk}$ cut of $0.97 \pm 0.05$, consistent with unity. The systematic error is obtained by varying the position and width of the vector boson mass peaks and the ratio of W to Z bosons.




(a) Leading jet mass for $n_{trk}$

(b) Leading jet mass for $n_{trk} \geq 30$

Figure 7.25: The leading jet mass distributions obtained by fitting a 4th order polynomial to all jets with $n_{trk} < 30$ (a) and $n_{trk} \geq 31$ (b) in the V+jets control region. The error band on the fit corresponds to the uncertainty on the mass scale. Similar fits are made in more granular bins of $n_{trk}$ and the $V + jet$ signal rate is extracted. Figure taken from [11].

## 7.7 Background Fit

As was briefly covered in the introduction, the analysis searches for a resonant peak on top of the smoothly falling background. Due to the low number of simulated background events with a large dijet mass passing the analysis selection, the background shape must be extracted from data. This process also allows the shape to account for minor contributions from $V+jets$ and $t\bar{t}$ events. The background shape is found to be well described by the parametric function:

$$\frac{dn}{dx} = p_1(1-x)^{p_2-\zeta p_3}x^{-p_3} \tag{7.3}$$

where $x = \sqrt{m_{JJ}}/\sqrt{s}$ and the function is fit to the number of events observed in each bin in $m_{JJ}$. Here $p_1$ is an overall normalisation factor and $p_2$ and $p_3$ are dimensionless shape parameters. $\zeta$ is fit simultaneously to minimise the correlations between $p_2$ and $p_3$. The fit is performed in the range $1.1 < m_{JJ} < 6.0$ TeV in bins of 100 GeV using a binned maximum-likelihood fit with likelihood function

$$\mathcal{L} = \prod_i \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!} \tag{7.4}$$

where $n_i$ is the number of events in the i$^{\text{th}}$ bin and $\lambda_i$ is the expected number of background events in that bin. $p_1, p_2$ and $p_3$ are floated in the fit and the preferred

| Sample | $\chi^2/nDOF$ | Probability | $p_2$ | $p_3$ | $\zeta$ |
|---|---|---|---|---|---|
| LowLow Sideband | 13.4/21 | 0.894 | $55.70 \pm 0.65$ | $5.68 \pm 0.52$ | 7.427 |
| HighLow Sideband | 19.23/18 | 0.379 | $58.11 \pm 0.62$ | $6.71 \pm 0.49$ | 7.433 |
| HighHigh Sideband | 7.0/15 | 0.958 | $58.77 \pm 1.16$ | $6.63 \pm 0.92$ | 7.473 |

Table 7.6: Background fit parameters for the mass sidebands.

value of $\zeta$ is derived iteratively.

This fit was tested in each of the mass and partial tag sidebands; the fitted parameters are given in table 7.6 and the distributions shown in figure 7.26.

## 7.8 Systematic Uncertainties

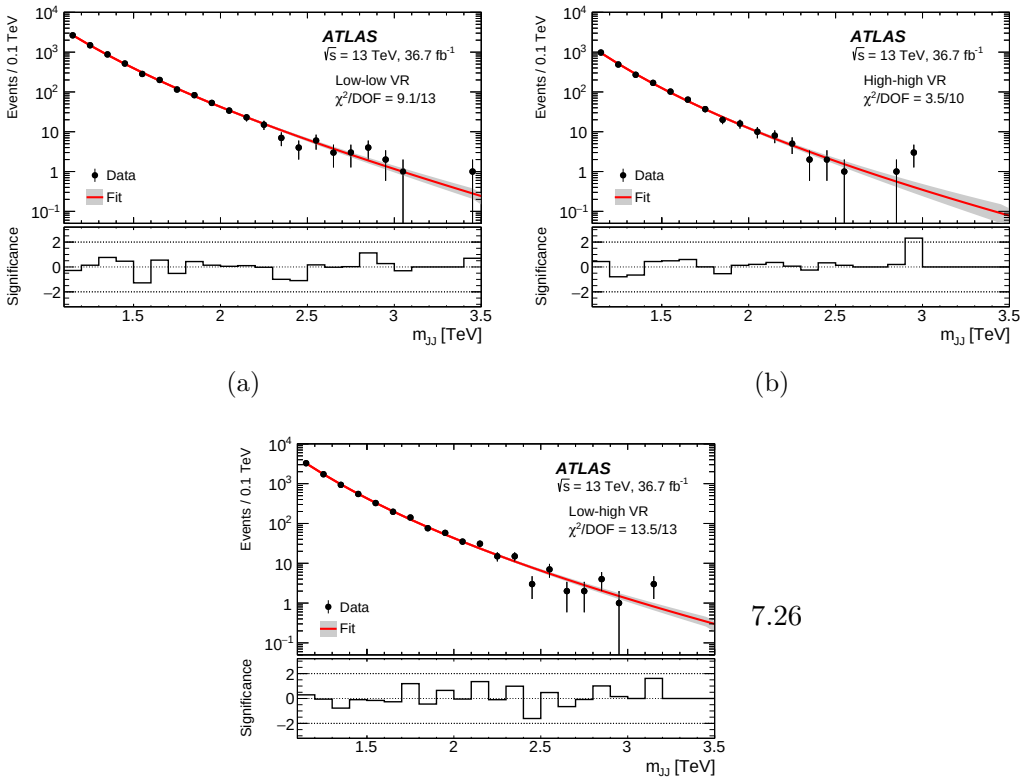
Systematic uncertainties on the analysis inputs are treated as nuisance parameters in the statistical analysis of the results. The dominant uncertainties on the signal yield and $m_{JJ}$ distribution shape are the uncertainty in the measurement of the large-R jet energy, mass, and $D_2$. The correlations between the scale uncertainties of these variables were considered, and the total uncertainty on the signal yield was calculated for three different cases: all three variables fully correlated, jet mass and energy correlated but $D_2$ uncorrelated, and all three variable uncorrelated. The partially correlated case was found to give the largest uncertainty, so this configuration was used.

The primary means for evaluating the jet uncertainties was by the so-called ‘‘double-ratio method’’ [78]. Take the ratio

$$r^X_{trk} = \frac{X_{calo}}{X_{trk}}, \tag{7.5}$$

the mean value of this quantity is expected agree between simulation and data provided that the detector simulation is correct. Any errors or omissions in the simulation of the jet production should cancel between the calorimeter and track terms when averaged over many jets, provided they affect both track and calorimeter jets equally. We take any deviation from 1 in the double ratio

$$R^X_{trk} = \frac{r^{X,data}_{trk}}{r^{X,MC}_{trk}} \tag{7.6}$$




7.26

Figure 7.26: The $m_{JJ}$ distributions for the (a) LowLow, (b) HighLow, and (c) HighHigh mass sidebands and their respective fits. The shaded bands show the uncertainty in the fit and the lower panels show the significance of the number of observed events in each bin with respect to the fit. Plots taken from [11].

as an estimate of the uncertainty of the simulation. This technique is used to evaluate the uncertainty on the Jet Energy Scale (JES), Jet Mass Scale (JMS) and the $D_2$ scale, which are found to be 5%, 6% and 5% respectively. The value for JES also includes contributions from tracking uncertainties and MC simulation taken from the Herwig/Pythia double-ratio on top of the base double-ratio measurement. Uncertainties in the JES shift the location of a signal peak over the background and hence they impact the significance of any measurement. The JMS and $D_2$ scale uncertainties affect the efficiency of signal selection and dijet mass shape.

The resolution (and uncertainty on this resolution) of energy, mass and $D_2$ measurements would have similar effects on the signal location and selection efficiency. The energy resolution was taken to be a fixed 2% as suggested by the ATLAS JetEtMiss working group, the mass and $D_2$ resolutions were taken as the width of gaussians fitted to each of their respective response functions in two fixed eta bins. The impact of energy, mass and $D_2$ resolution uncertainties on signal acceptance and measured signal width was then evaluated by applying the uncertainties as gaussian smear to the jet observables and rerunning the analysis.

The uncertainty on the $n_{trk}$ cut was evaluated as discussed in §7.6.4, giving an additional $3 \pm 6\%$ tracks per jet in data compared to simulation. This gives an overall uncertainty of 12% at the event-level.

Several further measurement and simulation uncertainties are considered. The PDF uncertainties can have a large effect on the expected signal production weight, and this uncertainty is evaluated using the method proposed by the PDF4LHC group, reweighting events to the CT14, MMHT2014, NNPDF3.0 and ATLAS-epWZ12 PDF sets then finding the central value and deviation of the $m_{JJ}$ distribution for the additional PDF sets. This is considered as an uncertainty on the signal acceptance rather than cross-section and is found to be 1% for graviton signals and to vary between 1-12% for HVT signals, depending on the signal mass. Uncertainties in the impact of ISR and FSR due to Monte-Carlo tune are evaluated by reweighting to five sets of systematic variations and found to contribute a 3% (5%) uncertainty to signal acceptance for the HVT (graviton) models.

The uncertainty on the luminosity measurement was evaluated using a methodology similar to that given in [160] and found to be 3.6% on the combined 2015+2016 dataset. The uncertainty on the trigger efficiency is considered to be negligible.

A summary of all uncertainties considered is given in table 7.7. It was found that the jet energy scale and jet mass scale had the largest impact on the analysis.

| Uncertainty | Value | Primary Source |
|---|---|---|
| Jet Energy Scale | 5% | $r^{trk}$ ratio |
| Jet Energy Resolution | 2% (absolute) | Cross-sample checks |
| Jet Mass Scale | 6% | $r^{trk}$ ratio |
| Jet Mass Resolution | 20% | Cross-sample checks |
| $D_2$ Scale | 5% | $r^{trk}$ ratio |
| $D_2$ Resolution | 15% | Cross-sample checks |
| $n_{trk}$ Cut Efficiency | 12% | §7.6.4 |
| PDF (Graviton) | 1% | Reweighting method |
| PDF (HVT) | 1-12% | Reweighting method |
| MC Tune (Graviton) | 5% | Reweighting method |
| MC Tune (HVT) | 3% | Reweighting method |
| Luminosity | 3.6% | [160] |
| Trigger Efficiency | - | §7.4 |

Table 7.7: A summary of the evaluated uncertainties on analysis inputs.

| Parameter | WZ | WW | ZZ |
|---|---|---|---|
| Observed Events | 904 | 497 | 618 |
| $\zeta$ | 7.5 | 7.3 | 7.5 |
| $p_2$ | $56.1 \pm 1.8$ | $50.9 \pm 2.1$ | $57.0 \pm 2.1$ |
| $p_3$ | $4.9 \pm 1.5$ | $6.0 \pm 1.5$ | $5.9 \pm 1.7$ |
| $\chi^2/NDF$ | 9.8/12 | 8.0/12 | 13.5/11 |

Table 7.8: The fit parameters and number of observed events for each signal region.

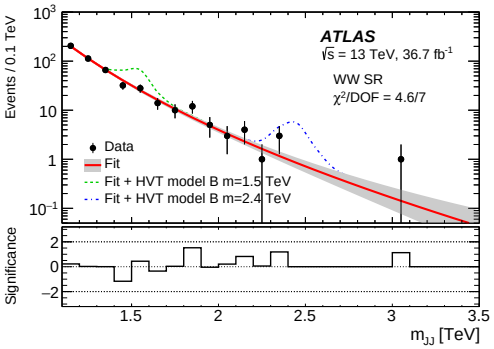
## 7.9 Results

### 7.9.1 Background Fit

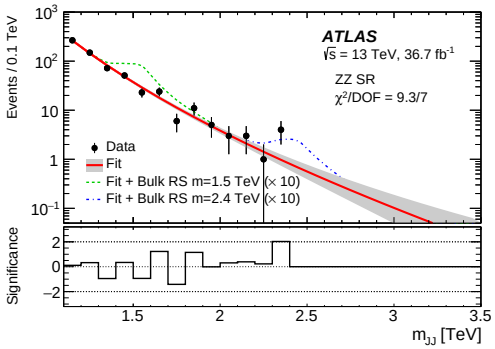
Background fits were made using events passing WW, WZ and ZZ selections as described in §7.7. The observed dijet mass spectra and fits are shown in figure 7.27. The fit parameters for each signal region are shown in table 7.8.

### 7.9.2 Statistical Analysis

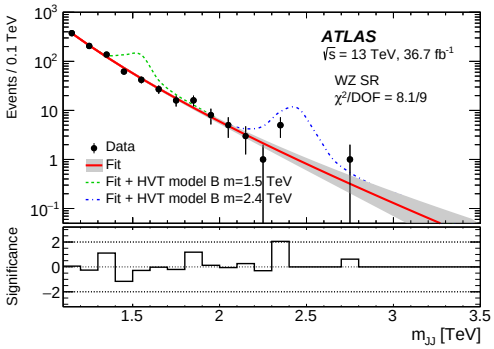
A frequentist statistical analysis is performed in each signal region for a range of signal mass points to evaluate the significance of any excesses and place limits on



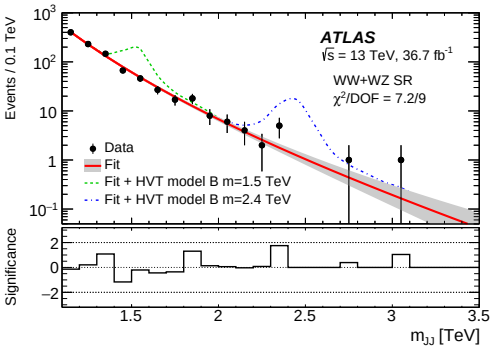

(a) WW Signal Region



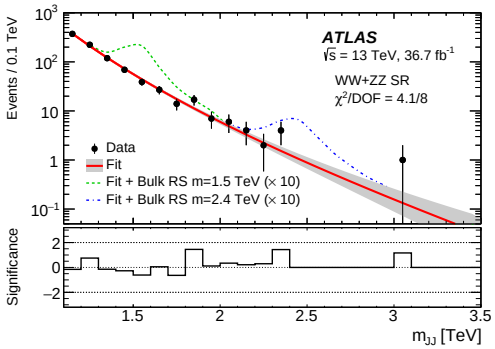

(b) ZZ Signal Region




(c) WZ Signal Region




(d) WZ+WW Signal Region




(e) WW+ZZ Signal Region

Figure 7.27: The fitted $m_{JJ}$ distributions for the $WW$ (a), $WZ$ (b), and $ZZ$ (c) signal regions. The combined WW+WZ (d) and WW+ZZ (e) signal regions are also shown. The background fit is shown by the red line and the shaded band corresponds to the uncertainty on this fit. The expected signal peaks are shown for HVT model B ($g_V = 3$) and Graviton models where appropriate. The graviton predictions are scaled by a factor of 10 for legibility.

| Model | Signal Region | Excluded mass range [GeV] |
|---|---|---|
| HVT model A, $g_V = 1$ | $WW$ | 1200 – 2200 |
| | $WZ$ | 1200 – 3000 |
| | $WW + WZ$ | 1200 – 3100 |
| HVT model B, $g_V = 3$ | $WW$ | 1200 – 2800 |
| | $WZ$ | 1200 – 3300 |
| | $WW + WZ$ | 1200 – 3500 |
| Bulk RS, $k/\bar{M}_{planck} = 1$ | $WW$ | 1300 – 1450 |
| | $ZZ$ | – |
| | $WW + ZZ$ | 1300 – 1600 |

Table 7.9: The mass regions for the HVT and RS graviton models that can be excluded at 95% confidence level by each individual and combined signal region. Table taken from [11].

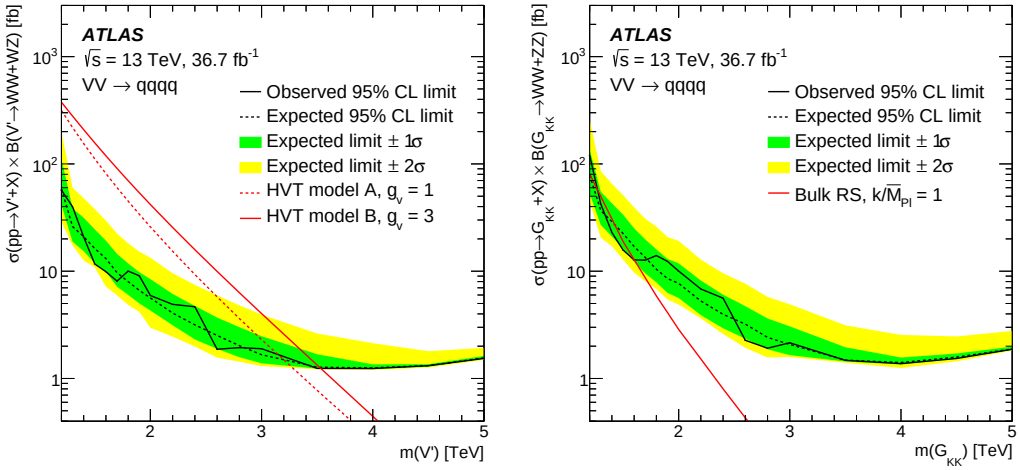
the $\sigma \times BR$. For each mass point, the parameter of interest measured is the signal strength $\mu$; $\mu = 0$ is the corresponds to the background-only hypothesis and $\mu = 1$ is the signal+background hypothesis. Information about $\mu$ is extracted using a test statistic $\lambda(\mu)$ based on a profile-likelihood ratio [161] where the likelihood is defined as

$$\mathcal{L} = \prod_i P_{pois}(n^i_{obs}|n^i_{exp}(\mu, p_1, p_2, p_3, \zeta)) \times G(\alpha) \times \mathcal{N}(\theta) \tag{7.7}$$

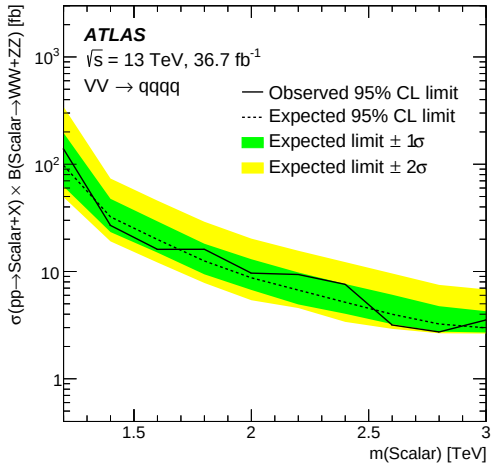
Here $P_{pois}(n^i_{obs}|n^i_{exp})$ is the poissonian probability of observing $n^i_{obs}$ events in the i$^{\text{th}}$ bin of $m_{JJ}$ given the expectation of $n^i_{exp}$ events in that bin (evaluated using eqn. (7.3) and signal expectation multiplied by $\mu$). $G(\alpha)$ is the set of gaussian PDFs corresponding to signal shape uncertainties and $\mathcal{N}(\theta)$ is the set of log-normal PDFs for nuisance parameters $\theta$ which model the uncertainty in the signal normalisation.

The local $p_0$ value, defined as the probability of the background-only model producing an excess at least as large as that observed, is used to quantify any excess found. This is translated into the global $p_0$ value of finding a similar excess across the whole search region using pseudo-experiments. The largest $p_0$ occurs in the WW channel for an HVT $Z'$ with a mass of 2 TeV and is approximately 6%.

Since no significant excess is observed, exclusion limits are set using the $CL_s$ method [162] at 95% confidence level, the regions excluded are given in table 7.9.




(a) $V' \to WW + WZ$ (b) $G \to WW + ZZ$




(c) $X \to WW + ZZ$

Figure 7.28: Upper-limits at the 95% $CL$ on the HVT (a) and (b) Bulk RS Graviton and heavy (c) scalar cross-section using the WW+WZ and WW+ZZ signal regions.

## 7.10 Conclusion

The analysis presented has updated the results given in [145] to include the complete $36.7fb^{-1}$ dataset collected by the ATLAS detector during 2015 and 2016 and made several significant improvements in methodology, including the use of the combined jet mass, re-derived boson tagging cuts and a reinterpretation in terms of a generic heavy scalar.

The results were consistent with the Standard Model-only hypothesis in all signal regions and limits were placed on HVT models A and B, the bulk RS graviton model and heavy scalar decays. A HVT model A (B) vector triplet is excluded with $g_V = 1$ (3) between 1.2 and 3.1 TeV (1.2 and 3.5 TeV), the RS graviton with $k/\overline{M}_{planck} = 1$ is excluded between 1.3 and 1.6 TeV and upper limits on $\sigma \times BR$ were placed on a heavy scalar, $X$, of 11 fb for $m_X = 2$ TeV and 3.9 fb for $m_X = 3$ TeV.

# Chapter 8

# Conclusion

Within the pages of this thesis I have attempted to give an overview of the varied work one must undertake when completing a Ph.D. as part of a particle physics experiment collaboration, and to provide enough background that a well-informed, or particularly dedicated, reader may engage with the remainder of the text.

The studies presented, multivariate calibration, ATLAS trigger performance, and a search for new particles, are obviously different in theme but are drawn together by a single thread linking all parts of this thesis - leveraging the substructure of large-R jets as a tool in the study of particle collisions.

§5 shows that jet substructure can deliver substantial improvements to the calibration of the large-R jet mass to the particle level by predicting the jet mass response using a neural network. A calibration of the jet energy was also attempted but was found to be less successful. Automatic variable selection methods were developed to prune the substantial library of jet substructure variables to a smaller neural network input set. Mutual information ranking and iterative principal component selection (IPCS) were proposed, and both were shown to select variables that contained useful information about the jet response. Deep and shallow neural networks were trained on MC QCD samples using these variable sets and a basic scan of the network architecture parameter space was undertaken to identify the best-performing energy and mass calibration networks.

After training, we identified the best-performing mass/energy calibration networks and input variable collections. For mass, this was a three layer 300-200-100 neuron structure trained using 10 variables selected using IPCS; for energy, the same network was used but only the energy, $p_T$ and $|\eta|$ were taken as inputs. The mass

calibration was found to give a 26% improvement in calibrated jet mass resolution compared to the current ATLAS ‘‘combined mass’’ calibration. The energy calibration was found to give no improvement. Further validation was performed on alternative generators, jet topologies, and on actual data, and the mass network was found to perform well in all cases. The data validation flagged up structures in the energy calibration network output caused by similar structures in the training sample. Further work is needed to identify how to remove these structures and to measure the mass calibration performance and uncertainties in data.

The following chapter, §6, contained two related studies on the ATLAS large-R jet trigger. The first used low energy small-R jet triggers to measure the efficiency of the large-R jet triggers used by ATLAS in 2015. It was shown that the standard online large-R jet reconstruction coupled with pileup subtraction and local cluster weighting of the jet constituents gave the best performance. The 360 GeV trigger was found to reach 100% efficiency by $p_T > 380$ GeV.

The second trigger study presented was more speculative, and showed that jet substructure could be used to create an online ‘‘anti-QCD’’ tagger that could be used to reject hadronic jets in favour of those from a variety of heavy object decays, which are the primary use-case of these jets in analyses. By combining the n-subjettiness variables several useful 1 vs 2 or 3 prong jet discriminants were created. Rough calculations indicate this could lower the jet trigger energy threshold to 349 GeV while maintaining the same rate as the current 395 GeV trigger. This tagger now needs to be implemented and tested within the ATLAS trigger software.

Finally, the primary physics analysis on which I worked during my PhD was presented in §7 [11]. This used large-R jets to search for heavy exotic particles decaying to vector bosons. These are expected from a variety of theoretical models, but the search was conducted in a relatively model-independent way, using the phenomenological framework of the Heavy Vector Triplet model as the primary analysis target. Both large-R jet calibrations and triggers were obviously important to this analysis, and the results shown in earlier chapters were either used in this analysis (in the case of the 2015 trigger studies) or would bring improvements to the analysis in a future iteration (in the case of the substructure trigger and calibration studies). Jet substructure also plays a significant role in this analysis as the basis of the boson-tagging techniques which make such a study possible. An important, though less glamorous, part of my work on this analysis was the development of the software for a large part of the analysis chain, so further details on this were provided.

The analysis identified collision events that had a topology consistent with a heavy object decaying to two highly boosted, hadronically-decaying, bosons. In the absence of new physics, selected events would primarily be caused by mis-tagged hadronic jets and would form a steeply falling background as a function of the mass of the dijet system. Any resonance would appear as a ''bump'' on this background. A smooth curve was fit to the mass spectrum observed, and a statistical analysis conducted to quantify any deviations from this fit at specific mass points. No significant deviations from the background were observed, and limits were placed on the HVT model A (B) and Graviton models of 1.2-3.1 TeV (1.2-3.5 TeV) and 1.3-1.6 TeV respectively.

I hope that the preceding thesis has left the reader with a deeper understanding of the intricacies of jets and the many ways that their substructure can contribute to physics at the LHC. Sadly, advances in analysis and reconstruction techniques such as these, and others used in the search presented, have not yet led to the discovery of new physics at the LHC. For now, whatever lies beyond the standard model is not yet obvious, but its many flaws seem to prove that something new must be out there.

# Appendix A

# Neural Network Calibration Results

## A.1 Training Results

This appendix contains the primary results for each calibration network generated as part of the study in §5. The tables are divided into the energy and mass calibration results and each shows the training and validation errors for each network, alongside the response and interquartile range measures for all of the datasets considered. Each table contains the training error, validation error, response, interquartile range and interdecile range for each network/input combination evaluated on each of the Pythia, Herwig, $WZ$, $t\bar{t}$ and $hh$ datasets. The best network in for each input set is highlighted in bold. Note that the energy IPCS networks were validated on a dataset that was missing the final 1/3 of jets, hence its validation error needs to be scaled by 3/2 to when compared to the other input sets.

| Energy - Simple Inputs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Pythia* | | | *Herwig* | | |
| Name | Train. Error | Valid. Error | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | - | - | 0.969 | 0.025 | 0.051 | 0.972 | 0.022 | 0.046 |
| Calib | - | - | 1.002 | 0.024 | 0.049 | 1.002 | 0.021 | 0.044 |
| S-1-50 | 2748.5 | 915.4 | 1.002 | 0.025 | 0.051 | 1.002 | 0.022 | 0.046 |
| S-1-100 | 2721.4 | 907.5 | 1.002 | 0.024 | 0.05 | 1.001 | 0.022 | 0.045 |
| S-1-200 | 2700.6 | 899.8 | 1.002 | 0.024 | 0.05 | 1.002 | 0.022 | 0.045 |
| S-2-50 | 2652.1 | 881.3 | 1.002 | 0.024 | 0.049 | 1.002 | 0.021 | 0.044 |
| S-2-100 | 2567.3 | 857.4 | 1.003 | 0.023 | 0.048 | 1.004 | 0.021 | 0.044 |
| S-3-50 | 2604.3 | 867.2 | 1.002 | 0.024 | 0.049 | 1.003 | 0.021 | 0.044 |
| S-3-100 | 2588.2 | 864.5 | 1.002 | 0.024 | 0.049 | 1.002 | 0.021 | 0.044 |
| **S-3-300** | **2544.7** | **851.8** | **1.002** | **0.023** | **0.048** | **1.002** | **0.021** | **0.043** |

| Energy - Simple Inputs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *WZ* | | | $t\bar{t}$ | | | *hh* | | |
| Name | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | 0.974 | 0.026 | 0.053 | 0.969 | 0.032 | 0.065 | 0.975 | 0.03 | 0.061 |
| Calib | 1.009 | 0.025 | 0.051 | 1.009 | 0.031 | 0.065 | 1.011 | 0.029 | 0.06 |
| S-1-50 | 1.009 | 0.026 | 0.053 | 1.008 | 0.032 | 0.065 | 1.011 | 0.03 | 0.062 |
| S-1-100 | 1.01 | 0.025 | 0.052 | 1.009 | 0.031 | 0.064 | 1.012 | 0.029 | 0.061 |
| S-1-200 | 1.01 | 0.025 | 0.052 | 1.01 | 0.031 | 0.064 | 1.013 | 0.029 | 0.06 |
| S-2-50 | 1.01 | 0.025 | 0.052 | 1.01 | 0.031 | 0.064 | 1.013 | 0.029 | 0.061 |
| S-2-100 | 1.011 | 0.025 | 0.051 | 1.011 | 0.031 | 0.064 | 1.013 | 0.029 | 0.061 |
| S-3-50 | 1.01 | 0.025 | 0.051 | 1.01 | 0.031 | 0.064 | 1.013 | 0.029 | 0.061 |
| S-3-100 | 1.01 | 0.025 | 0.051 | 1.01 | 0.031 | 0.064 | 1.013 | 0.029 | 0.06 |
| S-3-300 | 1.01 | 0.025 | 0.052 | 1.011 | 0.031 | 0.064 | 1.013 | 0.029 | 0.061 |

Table A.1: Energy calibration network results for the Simple input set.

| Energy - MI5 Inputs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Pythia* | | | *Herwig* | | |
| Name | Train. Error | Valid. Error | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | - | - | 0.969 | 0.025 | 0.051 | 0.972 | 0.022 | 0.046 |
| Calib | - | - | 1.002 | 0.024 | 0.049 | 1.002 | 0.021 | 0.044 |
| L-1-500 | 2649.7 | 887.2 | 1.005 | 0.024 | 0.05 | 1.005 | 0.022 | 0.045 |
| **L-2-100** | **2611.4** | **874.0** | **1.002** | **0.024** | **0.049** | **1.002** | **0.021** | **0.044** |
| L-3-100 | 2629.5 | 875.2 | 1.002 | 0.024 | 0.049 | 1.004 | 0.022 | 0.045 |
| L-3-300 | 2601.5 | 874.2 | 1.003 | 0.024 | 0.049 | 1.004 | 0.021 | 0.044 |
| L-4-100 | 2652.6 | 882.9 | 0.999 | 0.024 | 0.049 | 1.001 | 0.021 | 0.044 |
| L-4-300 | 2640.4 | 877.5 | 1.003 | 0.024 | 0.049 | 1.004 | 0.022 | 0.045 |
| L-5-100 | 2655.4 | 884.5 | 1.001 | 0.024 | 0.05 | 1.002 | 0.022 | 0.045 |
| L-5-300 | 2677.4 | 894.1 | 1.001 | 0.024 | 0.05 | 1.002 | 0.022 | 0.045 |

| Energy - MI5 Inputs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $WZ$ | | | $t\bar{t}$ | | | $hh$ | | |
| Name | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | 0.974 | 0.026 | 0.053 | 0.969 | 0.032 | 0.065 | 0.975 | 0.03 | 0.061 |
| Calib | 1.009 | 0.025 | 0.051 | 1.009 | 0.031 | 0.065 | 1.011 | 0.029 | 0.06 |
| L-1-500 | 1.014 | 0.025 | 0.052 | 1.013 | 0.031 | 0.065 | 1.016 | 0.029 | 0.061 |
| L-2-100 | 1.01 | 0.025 | 0.051 | 1.01 | 0.031 | 0.064 | 1.013 | 0.029 | 0.06 |
| L-3-100 | 1.008 | 0.025 | 0.051 | 1.008 | 0.031 | 0.064 | 1.011 | 0.029 | 0.06 |
| L-3-300 | 1.011 | 0.025 | 0.051 | 1.01 | 0.031 | 0.064 | 1.013 | 0.029 | 0.06 |
| L-4-100 | 1.006 | 0.025 | 0.051 | 1.006 | 0.031 | 0.064 | 1.009 | 0.029 | 0.06 |
| L-4-300 | 1.011 | 0.025 | 0.052 | 1.011 | 0.031 | 0.064 | 1.012 | 0.029 | 0.06 |
| L-5-100 | 1.008 | 0.025 | 0.052 | 1.006 | 0.031 | 0.064 | 1.009 | 0.029 | 0.06 |
| L-5-300 | 1.007 | 0.025 | 0.052 | 1.006 | 0.031 | 0.064 | 1.008 | 0.029 | 0.061 |

Table A.2: Energy calibration network results for the MI5 input set.

| Energy - MI10 Inputs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Pythia* | | | *Herwig* | | |
| Name | Train. Error | Valid. Error | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | - | - | 0.969 | 0.025 | 0.051 | 0.972 | 0.022 | 0.046 |
| Calib | - | - | 1.002 | 0.024 | 0.049 | 1.002 | 0.021 | 0.044 |
| **L-2-100** | **2667.1** | **886.4** | **1.001** | **0.024** | **0.049** | **1.003** | **0.022** | **0.045** |
| L-3-100 | 2668.3 | 886.9 | 1.001 | 0.024 | 0.05 | 1.002 | 0.022 | 0.045 |
| L-3-300 | 2668.8 | 895.3 | 1.004 | 0.026 | 0.053 | 1.006 | 0.023 | 0.048 |
| L-4-100 | 2989.0 | 992.4 | 1.007 | 0.024 | 0.05 | 1.009 | 0.022 | 0.045 |
| L-4-300 | 2766.5 | 929.6 | 1.002 | 0.024 | 0.049 | 1.002 | 0.021 | 0.045 |
| L-5-100 | 2663.2 | 892.0 | 1.0 | 0.025 | 0.051 | 1.001 | 0.022 | 0.046 |
| L-5-300 | 2760.7 | 923.2 | 1.004 | 0.024 | 0.051 | 1.005 | 0.022 | 0.046 |

| Energy - MI10 Inputs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $WZ$ | | | $t\bar{t}$ | | | $hh$ | | |
| Name | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | 0.974 | 0.026 | 0.053 | 0.969 | 0.032 | 0.065 | 0.975 | 0.03 | 0.061 |
| Calib | 1.009 | 0.025 | 0.051 | 1.009 | 0.031 | 0.065 | 1.011 | 0.029 | 0.06 |
| L-2-100 | 1.008 | 0.025 | 0.052 | 1.008 | 0.031 | 0.064 | 1.01 | 0.029 | 0.06 |
| L-3-100 | 1.007 | 0.025 | 0.052 | 1.005 | 0.031 | 0.064 | 1.008 | 0.029 | 0.061 |
| L-3-300 | 1.008 | 0.027 | 0.056 | 1.004 | 0.033 | 0.067 | 1.01 | 0.031 | 0.064 |
| L-4-100 | 1.014 | 0.025 | 0.053 | 1.012 | 0.031 | 0.065 | 1.016 | 0.03 | 0.061 |
| L-4-300 | 1.01 | 0.025 | 0.052 | 1.011 | 0.031 | 0.065 | 1.012 | 0.029 | 0.06 |
| L-5-100 | 1.006 | 0.025 | 0.053 | 1.004 | 0.031 | 0.065 | 1.008 | 0.03 | 0.061 |
| L-5-300 | 1.01 | 0.025 | 0.053 | 1.008 | 0.032 | 0.065 | 1.011 | 0.03 | 0.062 |

Table A.3: Energy calibration network results for the MI10 input set.

| Energy - IPCS Inputs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Pythia* | | | *Herwig* | | |
| Name | Train. Error | Valid. Error | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | - | - | 0.969 | 0.025 | 0.051 | 0.972 | 0.022 | 0.046 |
| Calib | - | - | 1.002 | 0.024 | 0.049 | 1.002 | 0.021 | 0.044 |
| L-1-500 | 1925.1 | 643.4 | 1.0 | 0.024 | 0.049 | 1.006 | 0.022 | 0.045 |
| **L-2-100** | **1871.7** | **627.1** | **1.003** | **0.023** | **0.048** | **1.01** | **0.021** | **0.044** |
| L-3-100 | 1890.0 | 628.1 | 1.0 | 0.023 | 0.048 | 1.003 | 0.021 | 0.045 |
| L-3-300 | 1951.0 | 648.8 | 1.001 | 0.024 | 0.049 | 1.005 | 0.022 | 0.045 |
| L-4-100 | 1886.3 | 629.1 | 1.002 | 0.023 | 0.048 | 1.006 | 0.021 | 0.044 |
| L-4-300 | 1997.1 | 666.4 | 1.003 | 0.024 | 0.05 | 1.004 | 0.022 | 0.045 |
| L-5-100 | 1995.4 | 663.5 | 1.002 | 0.024 | 0.05 | 1.006 | 0.022 | 0.045 |
| L-5-300 | 2744.7 | 918.3 | 1.003 | 0.024 | 0.051 | 1.009 | 0.023 | 0.047 |

| Energy - IPCS Inputs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *WZ* | | | $t\bar{t}$ | | | *hh* | | |
| Name | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | 0.974 | 0.026 | 0.053 | 0.969 | 0.032 | 0.065 | 0.975 | 0.03 | 0.061 |
| Calib | 1.009 | 0.025 | 0.051 | 1.009 | 0.031 | 0.065 | 1.011 | 0.029 | 0.06 |
| L-1-500 | 1.012 | 0.025 | 0.052 | 1.005 | 0.031 | 0.064 | 1.008 | 0.029 | 0.06 |
| L-2-100 | 1.016 | 0.025 | 0.052 | 1.013 | 0.031 | 0.065 | 1.017 | 0.03 | 0.061 |
| L-3-100 | 1.006 | 0.025 | 0.052 | 1.0 | 0.031 | 0.064 | 1.004 | 0.029 | 0.061 |
| L-3-300 | 1.01 | 0.025 | 0.052 | 1.006 | 0.031 | 0.063 | 1.009 | 0.029 | 0.06 |
| L-4-100 | 1.014 | 0.025 | 0.052 | 1.012 | 0.031 | 0.064 | 1.014 | 0.029 | 0.06 |
| L-4-300 | 1.01 | 0.025 | 0.052 | 1.009 | 0.031 | 0.064 | 1.012 | 0.029 | 0.06 |
| L-5-100 | 1.013 | 0.026 | 0.053 | 1.011 | 0.032 | 0.065 | 1.014 | 0.03 | 0.061 |
| L-5-300 | 1.019 | 0.027 | 0.056 | 1.02 | 0.034 | 0.068 | 1.026 | 0.032 | 0.065 |

Table A.4: Energy calibration network results for the IPCS input set.

| Mass - Simple Inputs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Pythia* | | | *Herwig* | | |
| Name | Train. Error | Valid. Error | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | - | - | 1.061 | 0.229 | 0.593 | 1.113 | 0.295 | 0.732 |
| Calib | - | - | 1.039 | 0.129 | 0.383 | 1.039 | 0.155 | 0.474 |
| S-1-50 | 1008210.0 | 336137.0 | 0.998 | 0.129 | 0.399 | 0.998 | 0.156 | 0.489 |
| S-1-100 | 1003890.0 | 333795.0 | 1.0 | 0.128 | 0.395 | 0.999 | 0.155 | 0.484 |
| S-1-200 | 1001760.0 | 334168.0 | 0.998 | 0.128 | 0.396 | 0.997 | 0.155 | 0.482 |
| S-2-50 | 1000890.0 | 333454.0 | 0.999 | 0.128 | 0.395 | 1.0 | 0.155 | 0.484 |
| S-2-100 | 996481.0 | 331884.0 | 0.999 | 0.127 | 0.392 | 0.998 | 0.154 | 0.482 |
| S-3-50 | 1002200.0 | 333963.0 | 0.999 | 0.128 | 0.396 | 1.001 | 0.156 | 0.486 |
| **S-3-100** | **996762.0** | **331554.0** | **0.996** | **0.126** | **0.386** | **0.994** | **0.151** | **0.474** |
| S-3-300 | 992154.0 | 332449.0 | 1.0 | 0.126 | 0.39 | 0.998 | 0.153 | 0.476 |

| Mass - Simple Inputs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *WZ* | | | $t\bar{t}$ | | | *hh* | | |
| Name | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | 1.006 | 0.083 | 0.18 | 0.965 | 0.059 | 0.13 | 0.975 | 0.062 | 0.13 |
| Calib | 1.024 | 0.072 | 0.153 | 1.021 | 0.064 | 0.141 | 1.035 | 0.064 | 0.14 |
| S-1-50 | 0.976 | 0.096 | 0.194 | 1.003 | 0.064 | 0.141 | 1.013 | 0.064 | 0.14 |
| S-1-100 | 0.98 | 0.097 | 0.198 | 1.005 | 0.063 | 0.14 | 1.015 | 0.064 | 0.14 |
| S-1-200 | 0.979 | 0.098 | 0.199 | 1.004 | 0.063 | 0.139 | 1.015 | 0.064 | 0.139 |
| S-2-50 | 0.978 | 0.09 | 0.196 | 0.99 | 0.06 | 0.131 | 1.0 | 0.063 | 0.135 |
| S-2-100 | 0.977 | 0.093 | 0.199 | 0.996 | 0.06 | 0.134 | 1.005 | 0.063 | 0.136 |
| S-3-50 | 0.972 | 0.089 | 0.193 | 0.988 | 0.061 | 0.133 | 0.996 | 0.063 | 0.135 |
| S-3-100 | 0.972 | 0.092 | 0.199 | 0.996 | 0.06 | 0.135 | 1.004 | 0.064 | 0.137 |
| S-3-300 | 0.98 | 0.094 | 0.201 | 1.001 | 0.061 | 0.135 | 1.011 | 0.064 | 0.138 |

Table A.5: Mass calibration network results for the Simple input set.

| Mass - TA Inputs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Pythia* | | | *Herwig* | | |
| Name | Train. Error | Valid. Error | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | - | - | 1.061 | 0.229 | 0.593 | 1.113 | 0.295 | 0.732 |
| Calib | - | - | 1.039 | 0.129 | 0.383 | 1.039 | 0.155 | 0.474 |
| S-1-50 | 964408.0 | 321651.0 | 0.997 | 0.122 | 0.37 | 0.998 | 0.147 | 0.456 |
| S-1-100 | 958722.0 | 319244.0 | 0.997 | 0.12 | 0.364 | 0.997 | 0.144 | 0.45 |
| S-1-200 | 956873.0 | 317696.0 | 1.001 | 0.119 | 0.364 | 1.001 | 0.144 | 0.448 |
| S-2-50 | 951727.0 | 316067.0 | 1.001 | 0.118 | 0.363 | 1.006 | 0.143 | 0.448 |
| **S-2-100** | **948190.0** | **314346.0** | **1.002** | **0.117** | **0.361** | **1.003** | **0.142** | **0.446** |
| S-3-50 | 951250.0 | 316434.0 | 1.004 | 0.118 | 0.367 | 1.008 | 0.145 | 0.454 |
| S-3-100 | 945607.0 | 316420.0 | 0.999 | 0.117 | 0.362 | 1.001 | 0.142 | 0.445 |
| S-3-300 | 941353.0 | 314820.0 | 1.0 | 0.116 | 0.357 | 1.0 | 0.14 | 0.44 |

| Mass - TA Inputs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $WZ$ | | | $t\bar{t}$ | | | $hh$ | | |
| Name | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | 1.006 | 0.083 | 0.18 | 0.965 | 0.059 | 0.13 | 0.975 | 0.062 | 0.13 |
| Calib | 1.024 | 0.072 | 0.153 | 1.021 | 0.064 | 0.141 | 1.035 | 0.064 | 0.14 |
| S-1-50 | 0.982 | 0.086 | 0.178 | 1.0 | 0.062 | 0.138 | 1.011 | 0.063 | 0.137 |
| S-1-100 | 0.982 | 0.087 | 0.179 | 1.003 | 0.062 | 0.138 | 1.013 | 0.062 | 0.137 |
| S-1-200 | 0.987 | 0.087 | 0.18 | 1.006 | 0.062 | 0.138 | 1.017 | 0.062 | 0.136 |
| S-2-50 | 0.98 | 0.081 | 0.176 | 0.991 | 0.059 | 0.128 | 1.003 | 0.061 | 0.132 |
| S-2-100 | 0.982 | 0.08 | 0.172 | 0.996 | 0.058 | 0.13 | 1.007 | 0.06 | 0.131 |
| S-3-50 | 0.974 | 0.081 | 0.171 | 0.991 | 0.06 | 0.129 | 1.0 | 0.062 | 0.132 |
| S-3-100 | 0.974 | 0.081 | 0.17 | 0.995 | 0.059 | 0.129 | 1.004 | 0.061 | 0.131 |
| S-3-300 | 0.981 | 0.081 | 0.173 | 1.001 | 0.059 | 0.132 | 1.011 | 0.06 | 0.132 |

Table A.6: Mass calibration network results for the TA input set.

| Mass - MI5 Inputs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Pythia* | | | *Herwig* | | |
| Name | Train. Error | Valid. Error | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | - | - | 1.061 | 0.229 | 0.593 | 1.113 | 0.295 | 0.732 |
| Calib | - | - | 1.039 | 0.129 | 0.383 | 1.039 | 0.155 | 0.474 |
| L-1-500 | 940772.0 | 313313.0 | 0.995 | 0.118 | 0.341 | 1.003 | 0.14 | 0.418 |
| L-2-100 | 942094.0 | 313377.0 | 1.001 | 0.121 | 0.352 | 1.012 | 0.145 | 0.432 |
| L-3-100 | 942253.0 | 315500.0 | 0.999 | 0.119 | 0.351 | 1.011 | 0.143 | 0.433 |
| **L-3-300** | **936060.0** | **311917.0** | **0.989** | **0.116** | **0.338** | **0.998** | **0.139** | **0.415** |
| L-4-100 | 953687.0 | 317323.0 | 0.979 | 0.118 | 0.348 | 0.989 | 0.143 | 0.43 |
| L-4-300 | 954849.0 | 317538.0 | 0.985 | 0.121 | 0.354 | 0.995 | 0.146 | 0.434 |
| L-5-100 | 971058.0 | 324414.0 | 0.955 | 0.117 | 0.336 | 0.964 | 0.137 | 0.42 |
| L-5-300 | 1182980.0 | 394045.0 | 0.874 | 0.127 | 0.315 | 0.876 | 0.143 | 0.399 |

| Mass - MI5 Inputs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *WZ* | | | $t\bar{t}$ | | | *hh* | | |
| Name | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | 1.006 | 0.083 | 0.18 | 0.965 | 0.059 | 0.13 | 0.975 | 0.062 | 0.13 |
| Calib | 1.024 | 0.072 | 0.153 | 1.021 | 0.064 | 0.141 | 1.035 | 0.064 | 0.14 |
| L-1-500 | 0.986 | 0.078 | 0.171 | 0.995 | 0.06 | 0.135 | 1.0 | 0.064 | 0.14 |
| L-2-100 | 0.992 | 0.079 | 0.169 | 0.988 | 0.061 | 0.133 | 0.995 | 0.063 | 0.139 |
| L-3-100 | 0.99 | 0.078 | 0.168 | 0.992 | 0.061 | 0.135 | 0.999 | 0.064 | 0.14 |
| L-3-300 | 0.98 | 0.076 | 0.167 | 0.989 | 0.061 | 0.134 | 0.994 | 0.063 | 0.139 |
| L-4-100 | 0.965 | 0.081 | 0.17 | 0.98 | 0.062 | 0.137 | 0.985 | 0.064 | 0.14 |
| L-4-300 | 0.968 | 0.079 | 0.168 | 0.98 | 0.063 | 0.137 | 0.985 | 0.065 | 0.141 |
| L-5-100 | 0.943 | 0.081 | 0.169 | 0.972 | 0.062 | 0.139 | 0.974 | 0.064 | 0.142 |
| L-5-300 | 0.858 | 0.091 | 0.18 | 0.932 | 0.073 | 0.154 | 0.925 | 0.075 | 0.159 |

Table A.7: Mass calibration network results for the MI5 input set.

| Mass - MI10 Inputs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Pythia* | | | *Herwig* | | |
| Name | Train. Error | Valid. Error | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | - | - | 1.061 | 0.229 | 0.593 | 1.113 | 0.295 | 0.732 |
| Calib | - | - | 1.039 | 0.129 | 0.383 | 1.039 | 0.155 | 0.474 |
| L-1-500 | 861465.0 | 287262.0 | 0.993 | 0.105 | 0.297 | 1.004 | 0.125 | 0.365 |
| L-2-100 | 848600.0 | 282569.0 | 1.001 | 0.103 | 0.299 | 1.02 | 0.126 | 0.373 |
| L-3-100 | 853982.0 | 285776.0 | 0.995 | 0.099 | 0.296 | 1.012 | 0.121 | 0.368 |
| **L-3-300** | **830209.0** | **277439.0** | **0.995** | **0.098** | **0.284** | **1.013** | **0.118** | **0.353** |
| L-4-100 | 860248.0 | 286416.0 | 1.001 | 0.103 | 0.305 | 1.017 | 0.127 | 0.382 |
| L-4-300 | 845213.0 | 281928.0 | 0.974 | 0.099 | 0.279 | 0.988 | 0.117 | 0.351 |
| L-5-100 | 853976.0 | 284656.0 | 0.995 | 0.101 | 0.297 | 1.013 | 0.123 | 0.373 |
| L-5-300 | 855001.0 | 284000.0 | 0.99 | 0.101 | 0.295 | 1.007 | 0.122 | 0.37 |

| Mass - MI10 Inputs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *WZ* | | | $t\bar{t}$ | | | *hh* | | |
| Name | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | 1.006 | 0.083 | 0.18 | 0.965 | 0.059 | 0.13 | 0.975 | 0.062 | 0.13 |
| Calib | 1.024 | 0.072 | 0.153 | 1.021 | 0.064 | 0.141 | 1.035 | 0.064 | 0.14 |
| L-1-500 | 0.997 | 0.068 | 0.142 | 0.995 | 0.059 | 0.129 | 1.003 | 0.059 | 0.128 |
| L-2-100 | 1.005 | 0.066 | 0.136 | 0.993 | 0.057 | 0.124 | 1.001 | 0.057 | 0.123 |
| L-3-100 | 0.995 | 0.065 | 0.134 | 0.99 | 0.058 | 0.122 | 0.998 | 0.057 | 0.123 |
| L-3-300 | 0.998 | 0.063 | 0.13 | 0.993 | 0.057 | 0.122 | 0.999 | 0.056 | 0.121 |
| L-4-100 | 0.991 | 0.064 | 0.134 | 0.993 | 0.058 | 0.123 | 0.999 | 0.057 | 0.124 |
| L-4-300 | 0.974 | 0.063 | 0.13 | 0.978 | 0.058 | 0.124 | 0.983 | 0.058 | 0.12 |
| L-5-100 | 0.997 | 0.067 | 0.14 | 0.992 | 0.059 | 0.125 | 1.0 | 0.059 | 0.126 |
| L-5-300 | 0.986 | 0.065 | 0.136 | 0.988 | 0.059 | 0.124 | 0.994 | 0.058 | 0.124 |

Table A.8: Mass calibration network results for the MI10 input set.

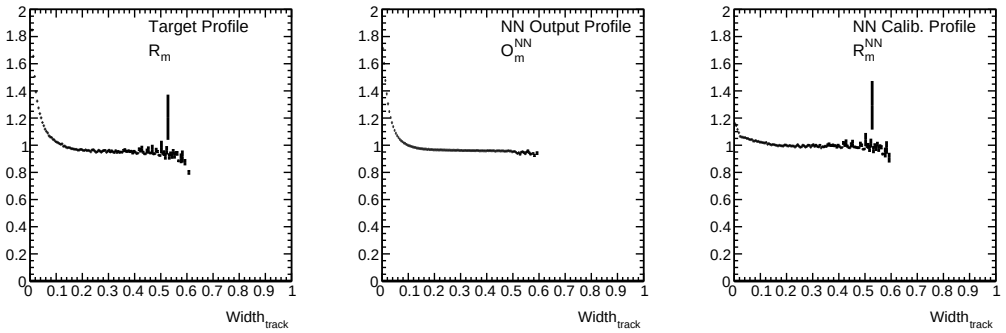
| Mass - IPCS Inputs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Pythia* | | | *Herwig* | | |
| Name | Train. Error | Valid. Error | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | - | - | 1.061 | 0.229 | 0.593 | 1.113 | 0.295 | 0.732 |
| Calib | - | - | 1.039 | 0.129 | 0.383 | 1.039 | 0.155 | 0.474 |
| L-1-500 | 801752.0 | 267260.0 | 0.989 | 0.096 | 0.264 | 0.991 | 0.112 | 0.321 |
| L-2-100 | 801316.0 | 266759.0 | 1.008 | 0.097 | 0.272 | 1.011 | 0.114 | 0.337 |
| L-3-100 | 811907.0 | 269774.0 | 1.004 | 0.098 | 0.279 | 1.008 | 0.116 | 0.347 |
| **L-3-300** | **799090.0** | **265911.0** | **0.997** | **0.095** | **0.274** | **1.001** | **0.113** | **0.341** |
| L-4-100 | 816304.0 | 272259.0 | 0.992 | 0.096 | 0.28 | 0.998 | 0.116 | 0.351 |
| L-4-300 | 806307.0 | 267844.0 | 0.994 | 0.095 | 0.271 | 0.999 | 0.112 | 0.337 |
| L-5-100 | 836070.0 | 277648.0 | 1.0 | 0.098 | 0.286 | 1.005 | 0.117 | 0.358 |
| L-5-300 | 825438.0 | 274746.0 | 0.985 | 0.098 | 0.282 | 0.989 | 0.116 | 0.357 |

| Mass - IPCS Inputs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *WZ* | | | $t\bar{t}$ | | | *hh* | | |
| Name | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ | $\hat{R}$ | $IQR(R)$ | $IDQ(R)$ |
| Uncalib | 1.006 | 0.083 | 0.18 | 0.965 | 0.059 | 0.13 | 0.975 | 0.062 | 0.13 |
| Calib | 1.024 | 0.072 | 0.153 | 1.021 | 0.064 | 0.141 | 1.035 | 0.064 | 0.14 |
| L-1-500 | 1.001 | 0.064 | 0.132 | 0.999 | 0.058 | 0.13 | 1.01 | 0.056 | 0.123 |
| L-2-100 | 1.016 | 0.065 | 0.135 | 1.01 | 0.057 | 0.128 | 1.021 | 0.058 | 0.121 |
| L-3-100 | 1.01 | 0.065 | 0.134 | 1.004 | 0.057 | 0.127 | 1.014 | 0.057 | 0.122 |
| L-3-300 | 1.002 | 0.062 | 0.129 | 0.995 | 0.056 | 0.123 | 1.008 | 0.056 | 0.121 |
| L-4-100 | 0.992 | 0.064 | 0.131 | 0.989 | 0.058 | 0.123 | 1.002 | 0.057 | 0.123 |
| L-4-300 | 0.999 | 0.064 | 0.131 | 0.993 | 0.056 | 0.122 | 1.003 | 0.056 | 0.121 |
| L-5-100 | 1.002 | 0.069 | 0.141 | 0.993 | 0.059 | 0.126 | 1.006 | 0.06 | 0.129 |
| L-5-300 | 0.989 | 0.067 | 0.141 | 0.987 | 0.059 | 0.125 | 1.0 | 0.059 | 0.126 |

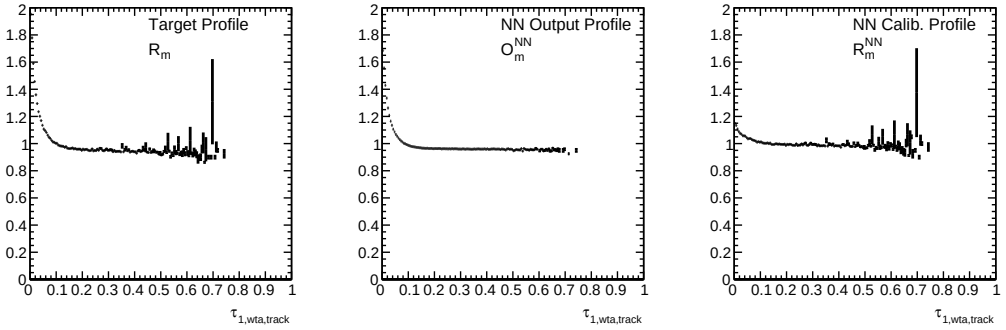
Table A.9: Mass calibration network results for the IPCS input set.

## A.2 Mass Network Profiling

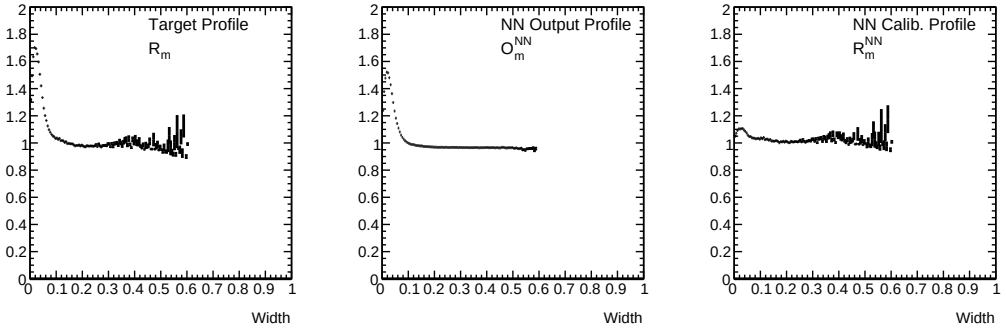
The following section contains the full set of profiling plots for the mass calibration network studied in §5.2.4.




(a) $W^{track}$

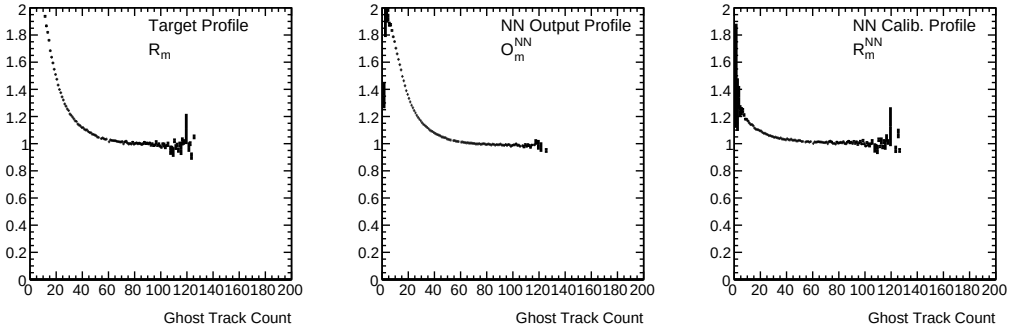



(b) $\tau_1^{track}$

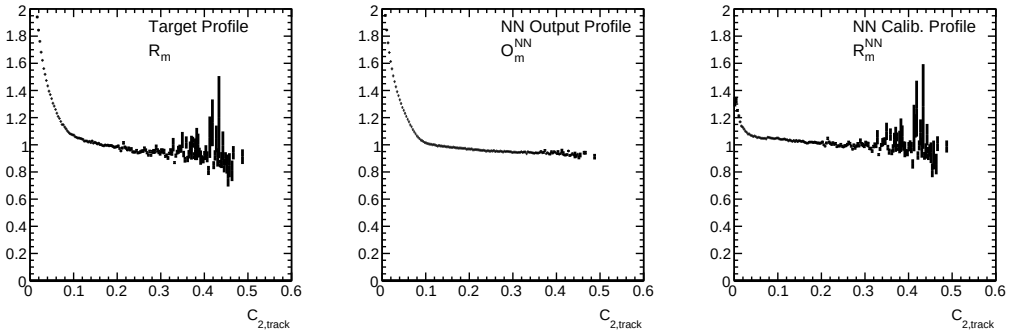



(c) $W$

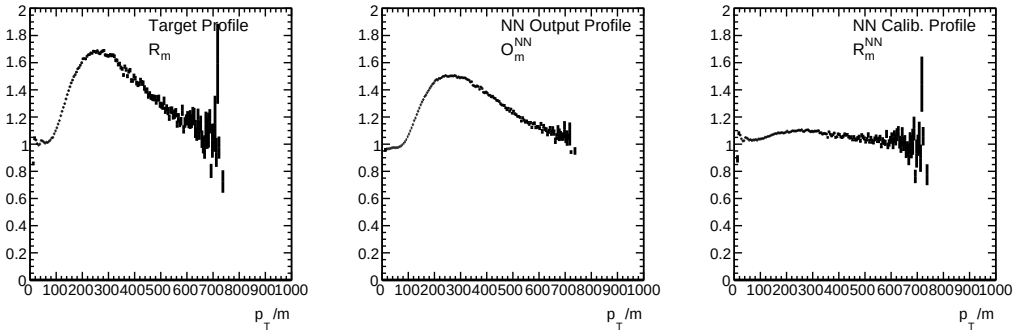
Figure A.1: Input response profiles for a subset of the input variables of the mass calibration network profiled in §5 [1/4].




(a) $n_{trk}$

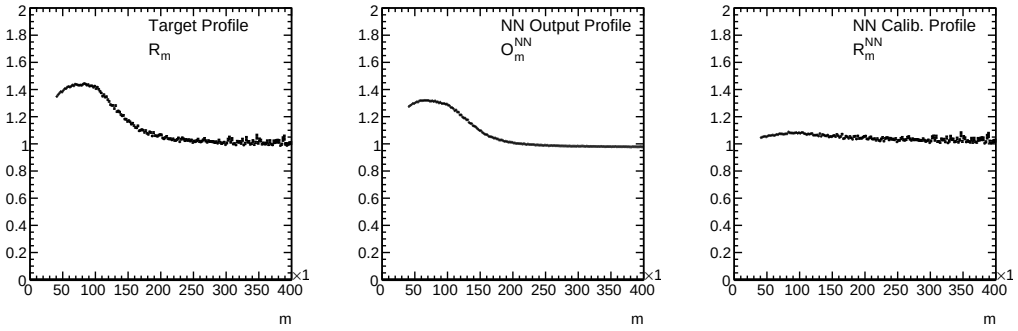



(b) $C_2^{track}$

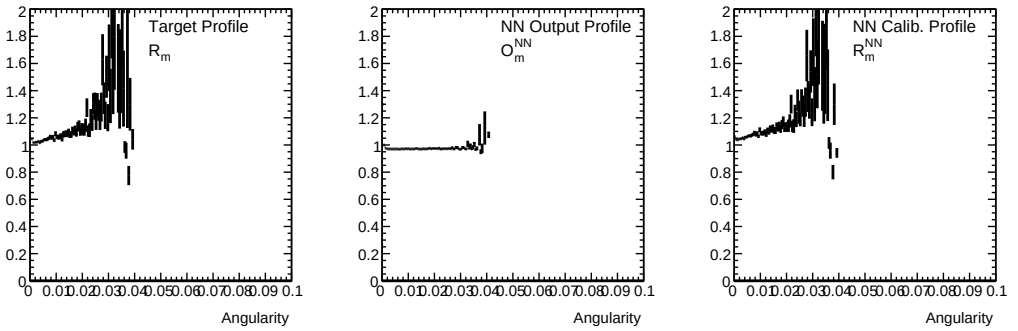



(c) $p_T/m$

Figure A.2: Input response profiles for a subset of the input variables of the mass calibration network profiled in §5 [2/4]

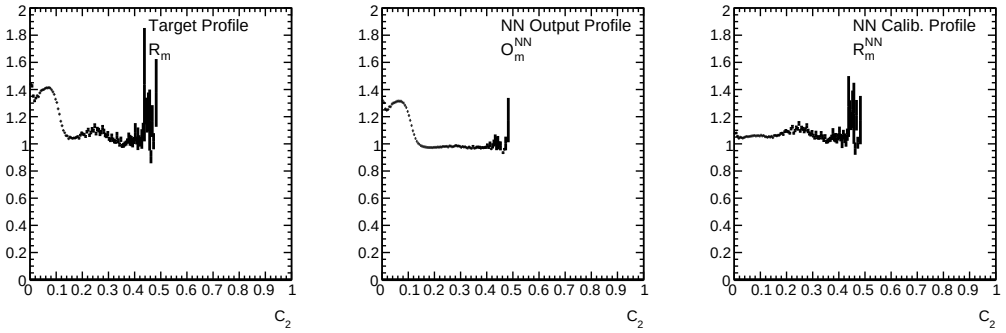



(a) $m$

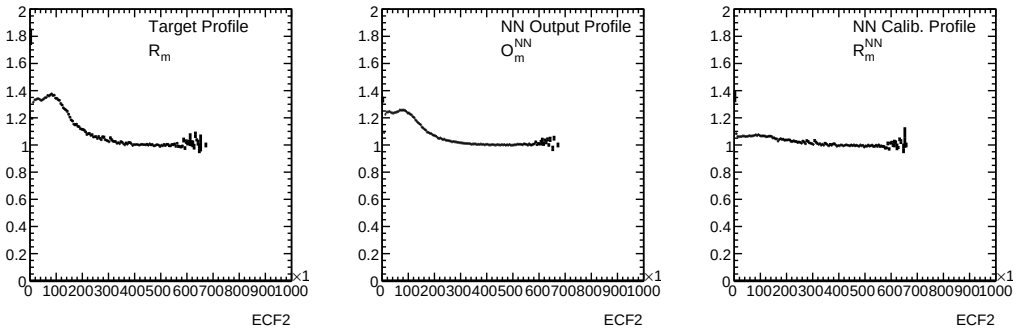



(b) $A$

Figure A.3: Input response profiles for a subset of the input variables of the mass calibration network profiled in §5 [3/4].

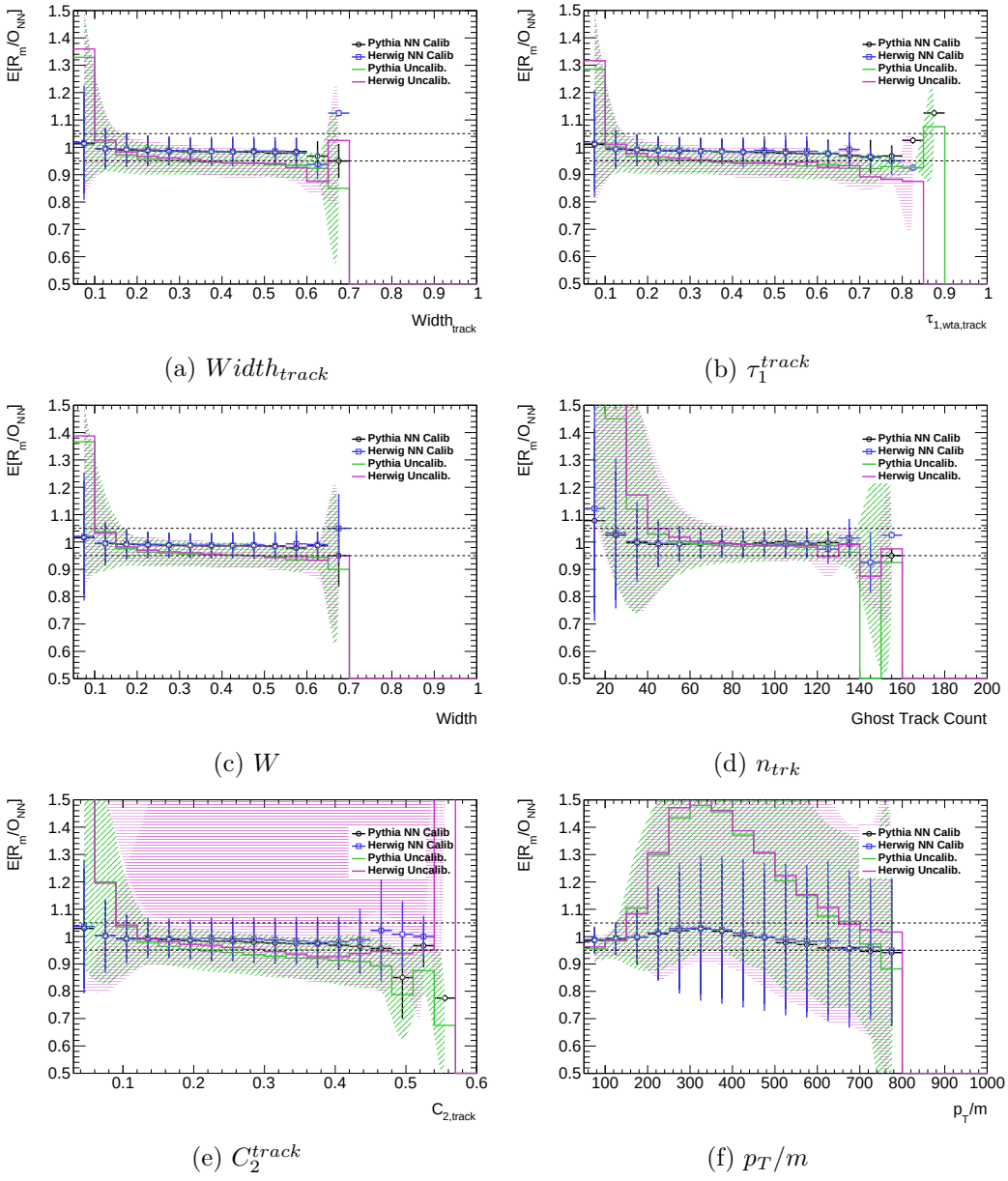



(a) $C_2$




(b) $ECF2$

Figure A.4: Input response profiles for a subset of the input variables of the mass calibration network profiled in §5 [4/4].

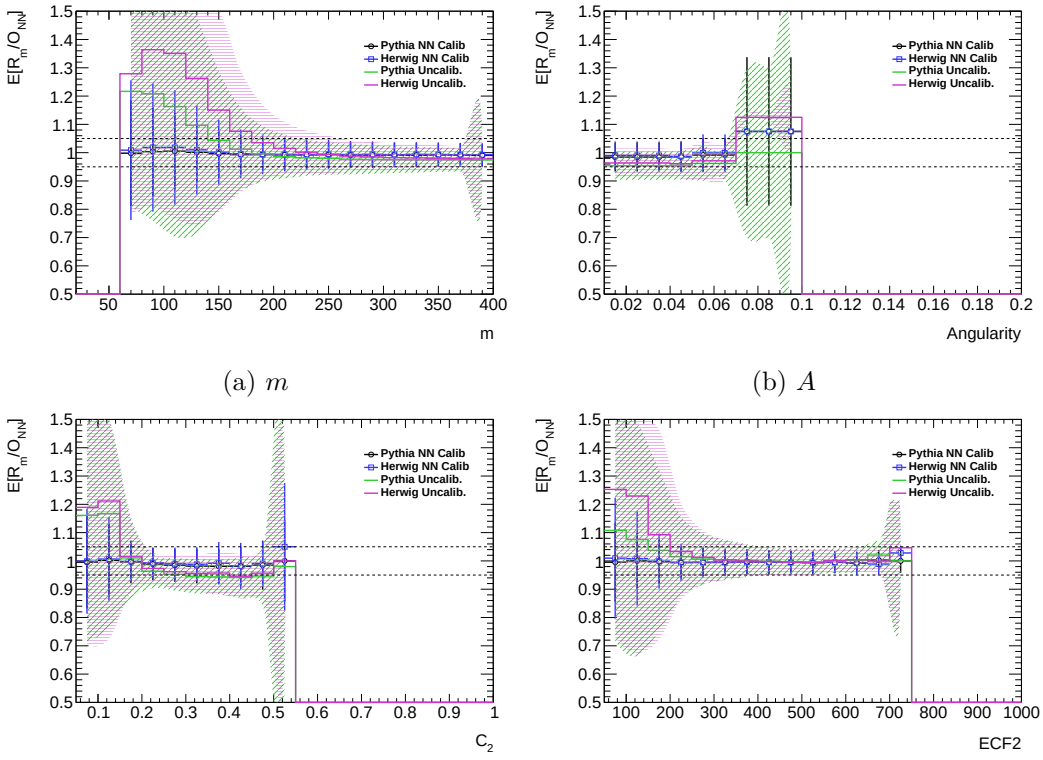



(a) $Width_{track}$ (b) $\tau_1^{track}$

(c) $W$ (d) $n_{trk}$

(e) $C_2^{track}$ (f) $p_T/m$

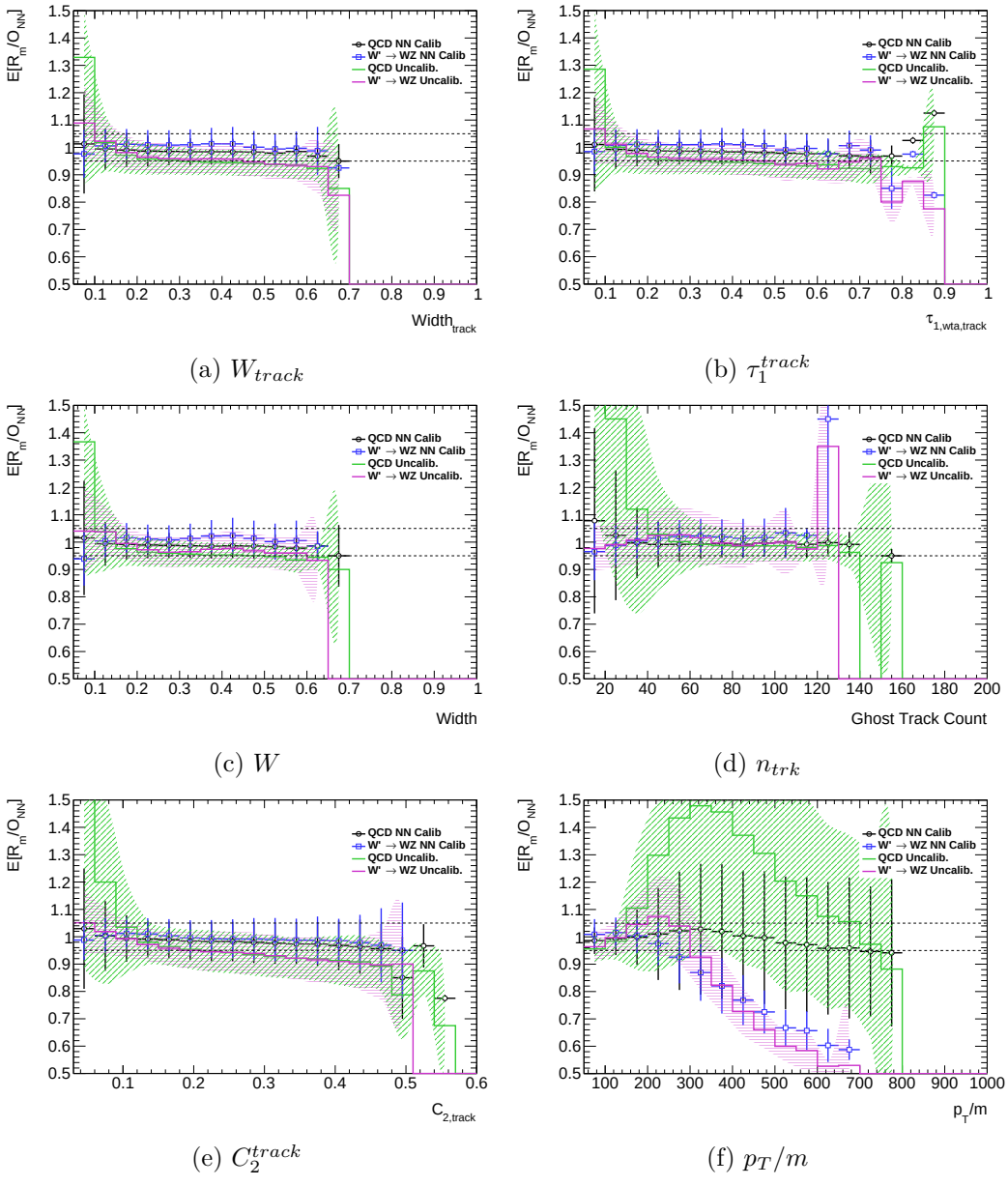
Figure A.5: A comparison of the closure for the mass calibration network when tested on jets from the Herwig QCD dataset and the Pythia QCD validation set [1/2].




(a) $m$ (b) $A$

(c) $C_2$ (d) $ECF_2$

Figure A.6: A comparison of the closure for the mass calibration network when tested on jets from the Herwig QCD dataset and the Pythia QCD validation set [2/2].

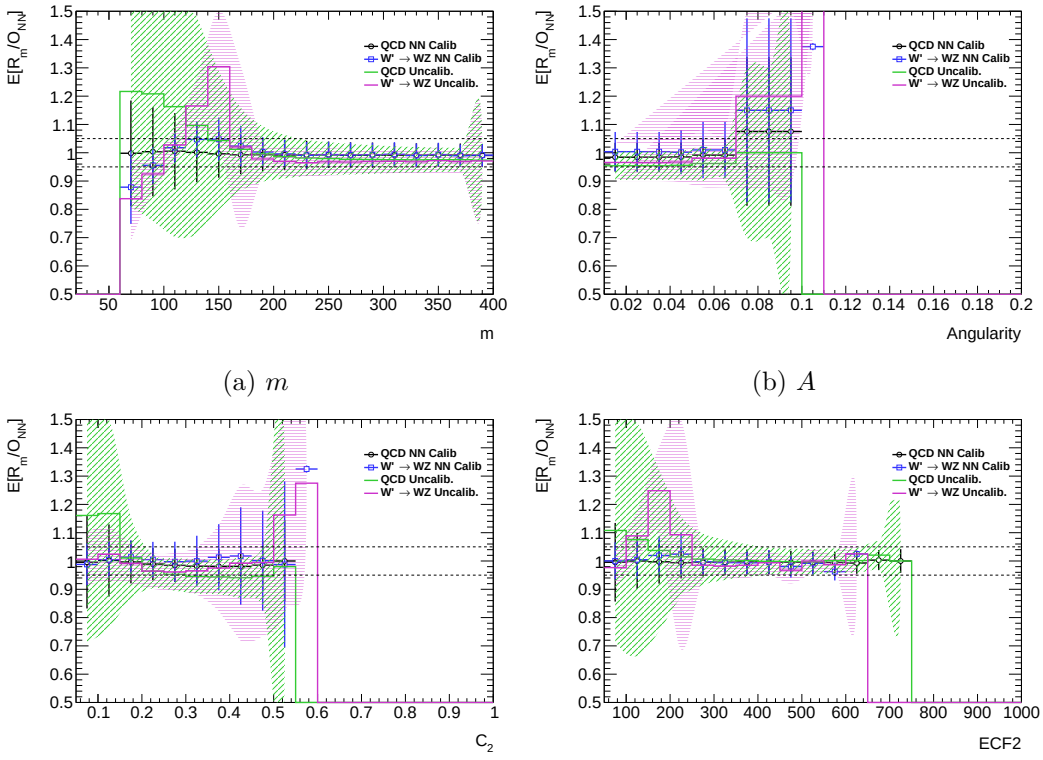



(a) $W_{track}$ (b) $\tau_1^{track}$

(c) $W$ (d) $n_{trk}$

(e) $C_2^{track}$ (f) $p_T/m$

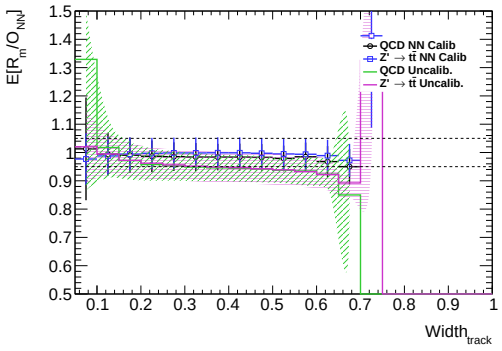
Figure A.7: A comparison of the closure for the mass calibration network when tested on jets from the $W' \to WZ$ dataset and the Pythia QCD validation set [1/2].



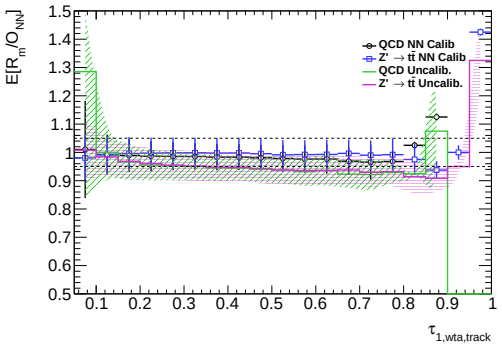

(a) $m$ (b) $A$

(c) $C_2$ (d) $ECF_2$

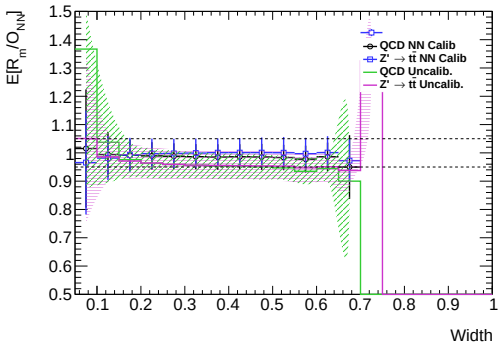
Figure A.8: A comparison of the closure for the mass calibration network when tested on jets from the $W' \to WZ$ dataset and the Pythia QCD validation set [2/2].



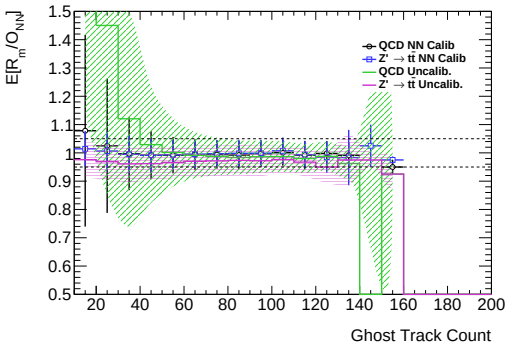

(a) $W_{track}$



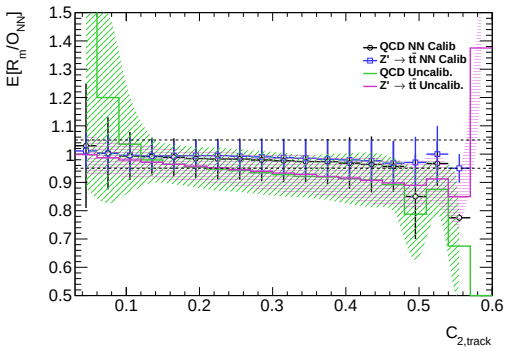

(b) $\tau_1^{track}$



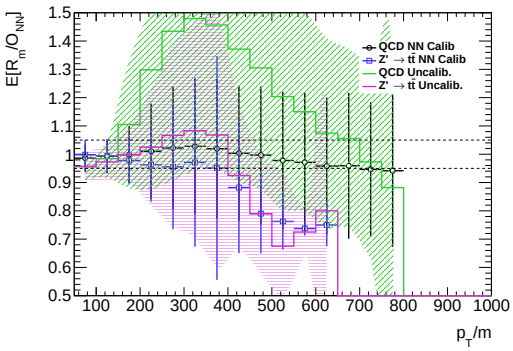

(c) $W$




(d) $n_{trk}$

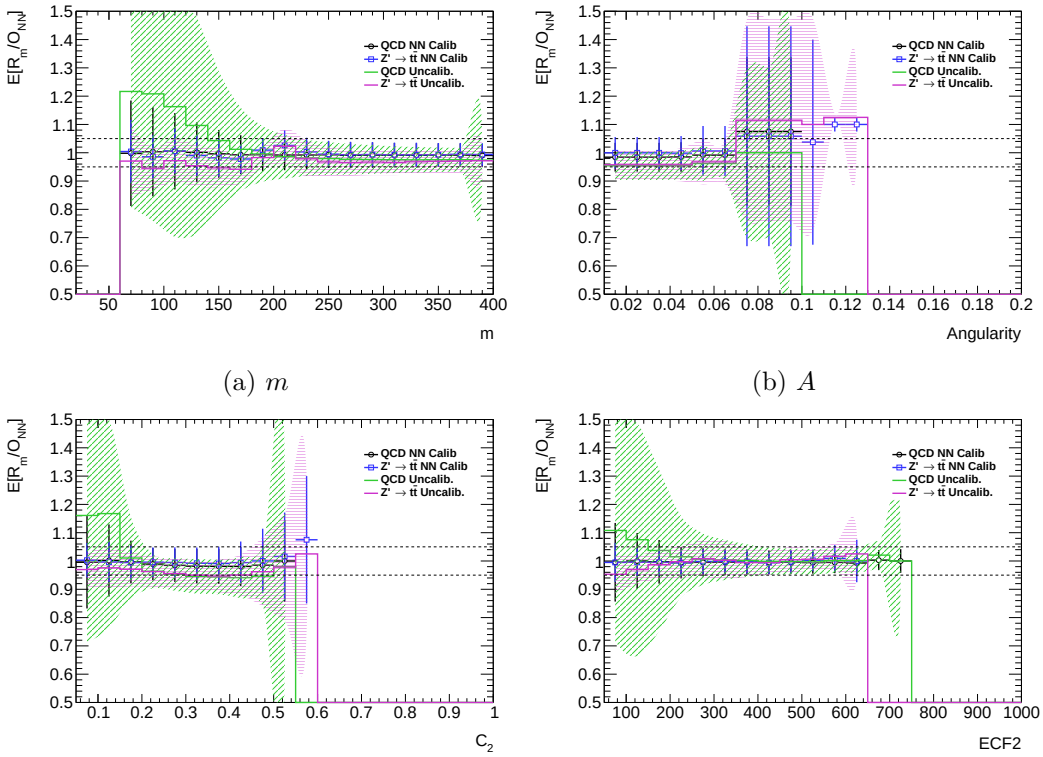



(e) $C_2^{track}$




(f) $p_T/m$

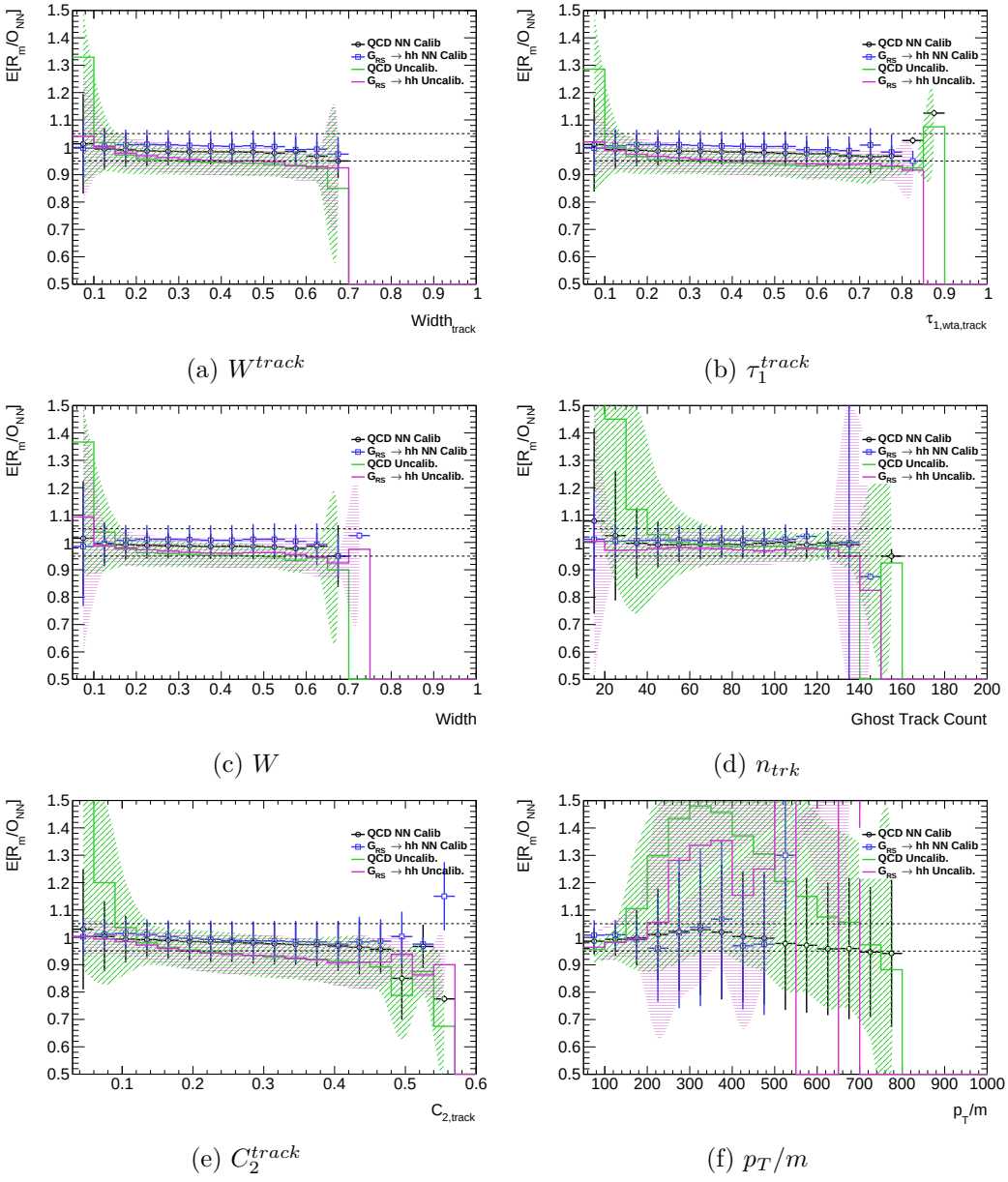
Figure A.9: A comparison of the closure for the mass calibration network when tested on jets from the $Z' \to t\bar{t}$ dataset and the Pythia QCD validation set [1/2].




(a) $m$ (b) $A$

(c) $C_2$ (d) $ECF_2$

Figure A.10: A comparison of the closure for the mass calibration network when tested on jets from the $Z' \to t\bar{t}$ dataset and the Pythia QCD validation set [2/2].

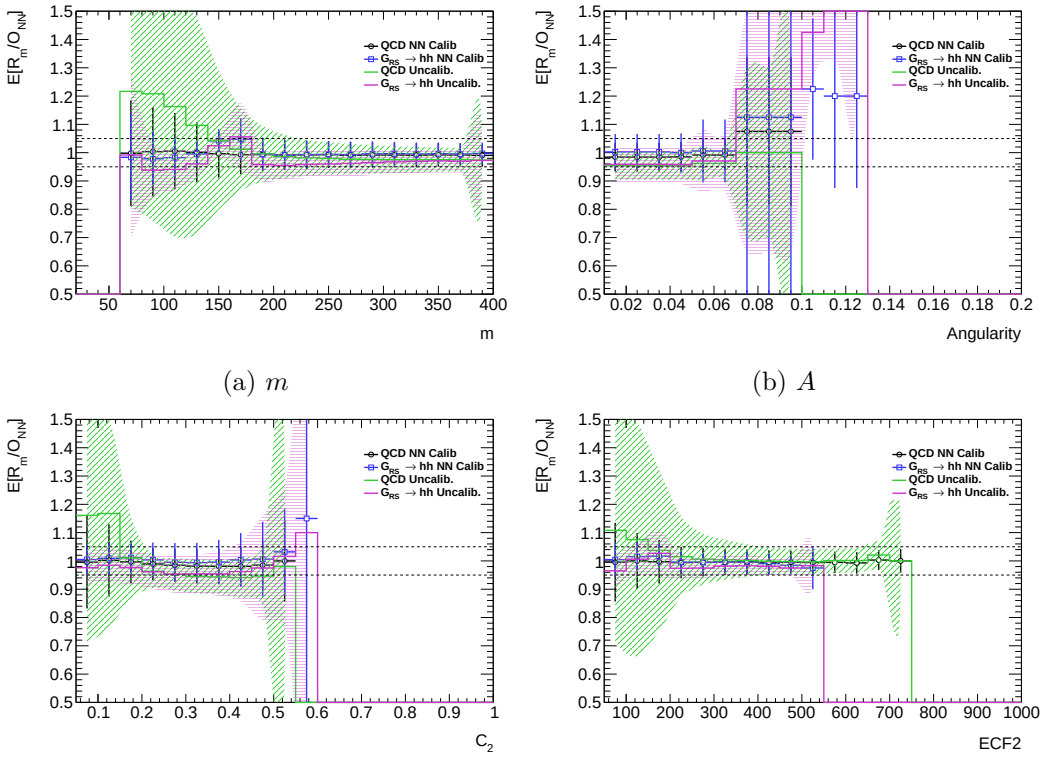



(a) $W^{track}$ (b) $\tau_1^{track}$

(c) $W$ (d) $n_{trk}$

(e) $C_2^{track}$ (f) $p_T/m$

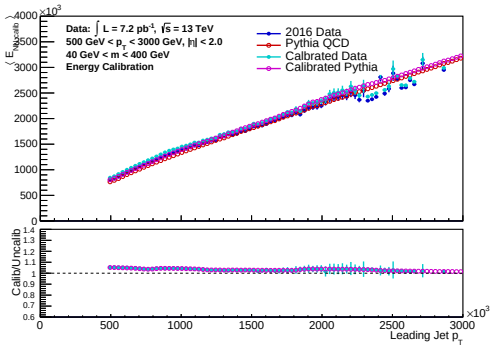
Figure A.11: A comparison of the closure for the mass calibration network when tested on jets from the $G_{RS} -> hh$ dataset and the Pythia QCD validation set [1/2].



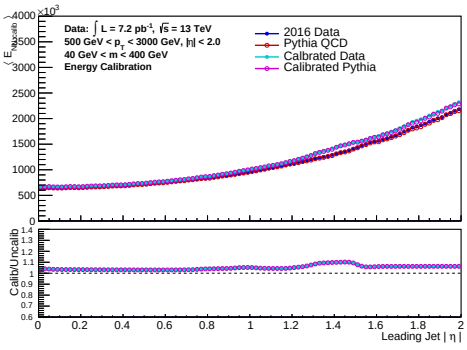

(a) $m$ (b) $A$

(c) $C_2$ (d) $ECF_2$

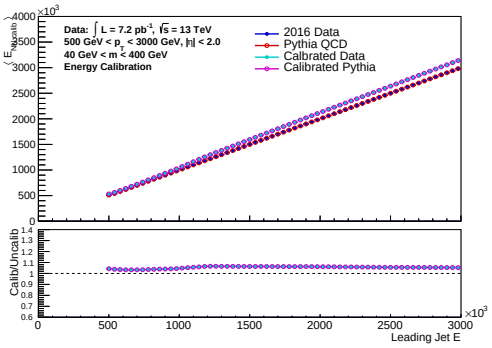
Figure A.12: A comparison of the closure for the mass calibration network when tested on jets from the $G_{RS} -> hh$ dataset and the Pythia QCD validation set [2/2].




(a) $p_T$




(b) $|\eta|$

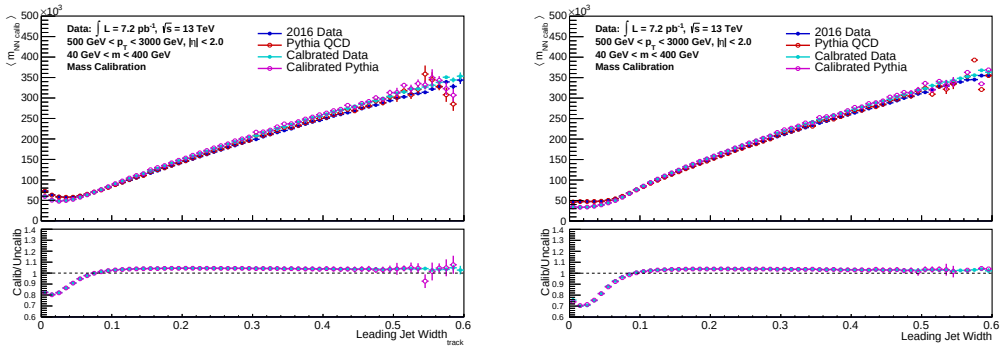



(c) $E$

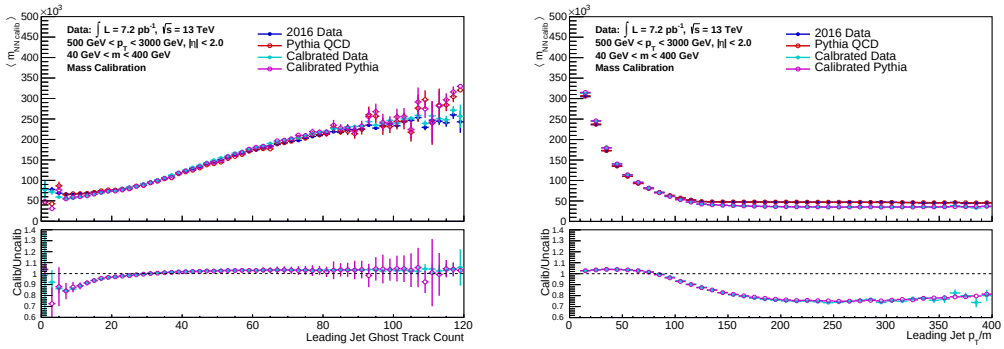
Figure A.13: The mean jet energy as a function of some of the energy calibration network's inputs, $p_T$, $|\eta|$, and $E$, before and after the calibration is applied. The ratios of the calibrated to uncalibrated distributions is shown below each plot for both data and MC samples. No unexpected shaping of the mass distribution is seen and no significant difference in behaviour is observed between data and MC.

## A.3 Calibrated Target Resolutions in Data

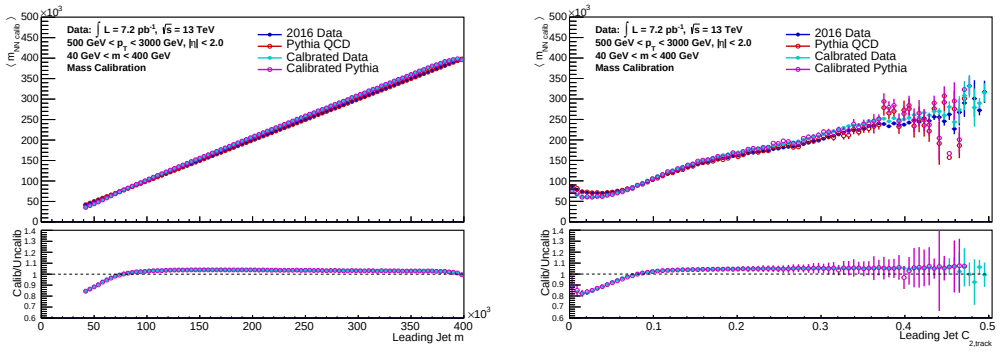
The following show the pre and post-calibration distributions of energy and mass generated by applying the training networks as a function of the network inputs. They use the data and monte-carlo samples prepared for the in-situ data validation of the network detailed in §5.2.2.




(a) $Width^{trk}$ (b) $Width$

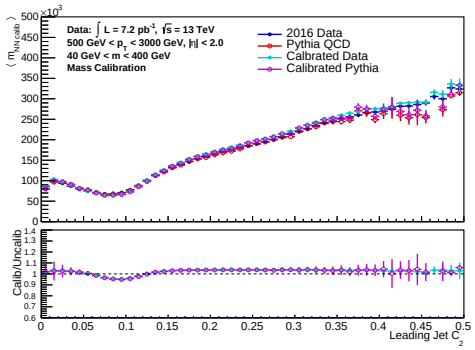



(c) $n^{trk}$ (d) $p_T/m$

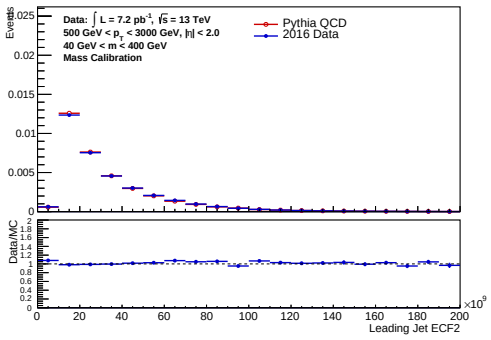



(e) $m$ (f) $C_2^{trk}$

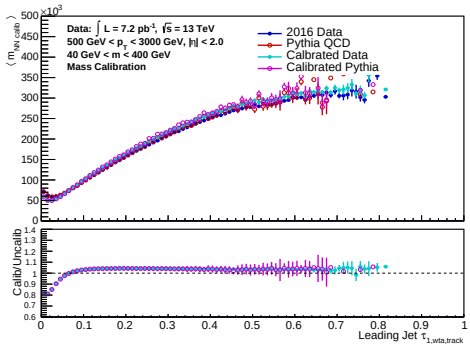
Figure A.14: The mean jet mass as a function of some of the mass calibration network's inputs, $Width^{trk}$, $Width$, $n_{trk}$, $p_T/m$, $m$, and $C_2^{trk}$, before and after the calibration is applied. The ratios of the calibrated to uncalibrated distributions is shown below each plot for both data and MC samples. No unexpected shaping of the mass distribution is seen and no significant difference in behaviour is observed between data and MC. The remaining input variables are shown in figure A.15.



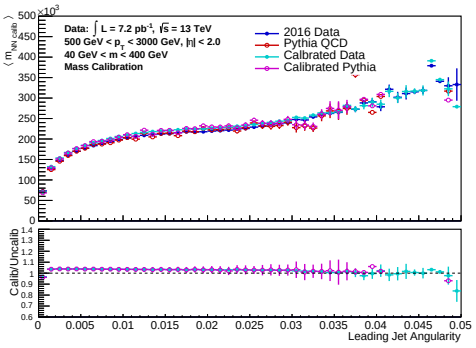

(a) $C_2$




(b) $ECF_2$




(c) $\tau_{21,wta}$




(d) $A$

Figure A.15: The mean jet mass as a function of some of the mass calibration network's inputs, $C_2$, $ECF_2$, $\tau^{trk}_{21,wta}$, and $A$, before and after the calibration is applied. The ratios of the calibrated to uncalibrated distributions is shown below each plot for both data and MC samples. No unexpected shaping of the mass distribution is seen and no significant difference in behaviour is observed between data and MC. The remaining variables are shown in figure A.14.

# Bibliography


[1] Douglas Adams. *The Restaurant at the End of the Universe*. Pan, 2009. ISBN 0330508598.

[2] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 GeV/$c^2$ at the CERN SPS collider. *Phys. Lett. B*, 126(5):398–410, Jun 1983. ISSN 03702693. doi: 10.1016/0370-2693(83)90188-0.

[3] D0 Collaboration. Observation of the Top Quark. *Phys. Rev. Lett.*, 74(14): 2632–2637, Apr 1995. doi: 10.1103/PhysRevLett.74.2632.

[4] CDF Collaboration. Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 74(14):2626–2631, Apr 1995. doi: 10.1103/PhysRevLett.74.2626.

[5] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716(1):1–29, 2012. ISSN 03702693. doi: 10.1016/j.physletb.2012.08.020.

[6] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716(1):30–61, 2012. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2012.08.021.

[7] Howard E Haber and Santa Cruz. Supersymmetry, Part 1 (Theory). Technical report, Particle Data Group (PDG), 2013.

[8] R.S. Chivukula, M. Narain, and J Womersley. Dynamical Electroweak Symmetry Breaking: Implications for the $H^0$. Technical report, Particle Data Group (PDG), 2015.

[9] S. Rolli and M. Tanabashi. Leptoquarks. Technical report, Particle Data Group (PDG).

[10] LHCb Collaboration. Test of lepton universality with $B_0 \to K_0^* \ell^+ \ell$ decays. *J. High Energy Phys.*, 2017(8):55, Aug 2017. ISSN 1029-8479. doi: 10.1007/JHEP08(2017)055.

[11] ATLAS Collaboration. Search for diboson resonances with boson-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B*, 777: 91–113, Feb 2018. doi: 10.1016/J.PHYSLETB.2017.12.011.

[12] Search for diboson resonance production in the $\ell\nu qq$ final state using $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC. Technical Report ATLAS-CONF-2016-062, CERN, Geneva, Aug 2016.

[13] ATLAS Collaboration. Searches for heavy diboson resonances in pp collisions at √s=13 TeV with the ATLAS detector. *J. High Energy Phys.*, 2016(9):173, Sep 2016. ISSN 10298479. doi: 10.1007/JHEP09(2016)173.

[14] Searches for heavy ZZ and ZW resonances in the llqq and vvqq final states in pp collisions at sqrt(s) = 13 TeV with the ATLAS detector. Technical Report ATLAS-CONF-2016-082, CERN, Geneva, Aug 2016.

[15] CMS Collaboration. Search for massive WH resonances decaying into the $\ell\nu bb$ final state at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, 76(5), May 2016. ISSN 14346052. doi: 10.1140/epjc/s10052-016-4067-z.

[16] CMS Collaboration. Search for Evidence of the Type-III Seesaw Mechanism in Multilepton Final States in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 119(22), Dec 2017. ISSN 10797114. doi: 10.1103/PhysRevLett.119.221802.

[17] ATLAS Collaboration. Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at √s = 8 TeV with the ATLAS detector. *J. High Energy Phys.*, 2015(12):1–39, Jun 2015. ISSN 10298479. doi: 10.1007/JHEP12(2015)055.

[18] Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the qq$^-$()bb$^-$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physics Letters B*, 774:494 – 515, 2017. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2017.09.066.

[19] ATLAS Collaboration. Performance of the ATLAS trigger system in 2010. *Eur. Phys. J. C*, 72(1):1–61, Nov 2012. ISSN 1434-6044. doi: 10.1140/epjc/s10052-011-1849-1.

[20] Zach Wienersmith. Saturday Morning Breakfast Cereal - To The Collider!, 2014.

[21] Lyndon Evans. The Large Hadron Collider. *Phil. Trans. R. Soc. A*, 370: 831–858, 2012. doi: 10.1098/rsta.2011.0453.

[22] et al L. Arnaudon. Linac4 Technical Design Report. Technical Report CERN-AB-2006-084. CARE-Note-2006-022-HIPPI, CERN, Geneva, Dec 2006. revised version submitted on 2006-12-14 09:00:40.

[23] Christiane Lefèvre. The CERN accelerator complex. Complexe des accélérateurs du CERN. Dec 2008.

[24] Jean-Luc Caron. Cross section of LHC dipole.. Dipole LHC: coupe transversale. AC Collection. Legacy of AC. Pictures from 1992 to 2002., May 1998.

[25] LHC Team. LHC Report: end of 2016 proton-proton operation | CERN, 2016.

[26] *ATLAS detector and physics performance: Technical Design Report, 2.* Technical Design Report ATLAS. CERN, Geneva, 1999.

[27] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *J. Instrum.*, 3(08):S08003, 2008.

[28] Alexander V. Zlobin. Superconducting Magnets - Principles, Operation, and Applications. In *Wiley Encycl. Electr. Electron. Eng.*, pages 1–19. John Wiley & Sons, Inc., Hoboken, NJ, USA, Dec 2014. doi: 10.1002/047134608X.W4520.pub2.

[29] M Capeans, G Darbo, K Einsweiller, M Elsing, T Flick, M Garcia-Sciveres, C Gemme, H Pernegger, O Rohne, and R Vuillermet. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010.

[30] ATLAS Collaboration. Performance of the ATLAS Inner Detector. Aug 2010.

[31] Randall Munroe. xkcd: Fundamental Forces, 2015.

[32] Ian Johnston Rhind Aitchison and Anthony J. G. Hey. *Gauge theories in particle physics : a practical introduction.* Institute of Physics Pub, 2003.

[33] D. Peskin, M. and Schroeder. *An Introduction to Quantum Field Theory.* Westview Press, 1995. ISBN 0201503972.

[34] Zee. *Quantum Field Theory in a Nutshell*, volume 1. Princeton University Press, 2015. ISBN 9788578110796. doi: 10.1017/CBO9781107415324.004.

[35] Peter Skands. Introduction to QCD. (arXiv:1207.2389. CERN-PH-TH-2012-196):341–420. 76 p, Jul 2012. Comments: Lecture notes from a course given at TASI 2012. Last update: January 2015. 79 pages, including a hyperlinked index at the back.

[36] Heather E Logan. TASI 2013 lectures on Higgs physics within and beyond the Standard Model. 2014.

[37] ATLAS Collaboration. Summary plots from the ATLAS Standard Model physics group, 2017.

[38] ATLAS Collaboration. Summary plots from the ATLAS SUSY physics group, 2014.

[39] J. D. Jackson and L. B. Okun. Historical roots of gauge invariance. *Rev. Mod. Phys.*, 73(3):663–680, Sep 2001. ISSN 00346861. doi: 10.1103/RevModPhys.73.663.

[40] Max Lautsch. *Running Couplings in the Standard Model and their Implications to possible Physics beyond the Standard Model*. PhD thesis, Heidelberg U., 2014.

[41] Luminita N. Mihaila, Jens Salomon, and Matthias Steinhauser. Renormalization constants and beta functions for the gauge couplings of the standard model to three-loop order. *Phys. Rev. D - Part. Fields, Gravit. Cosmol.*, 86 (9):96008, Nov 2012. ISSN 15507998. doi: 10.1103/PhysRevD.86.096008.

[42] CMS Collaboration. Summaries of CMS cross section measurements, 2017.

[43] Measurement of multi-jet cross-section ratios and determination of the strong coupling constant in proton-proton collisions at sqrt(s)=7 TeV with the ATLAS detector. Technical Report ATLAS-CONF-2013-041, CERN, Geneva, Apr 2013.

[44] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C.-P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93(3):033006, Feb 2016. ISSN 2470-0010. doi: 10.1103/PhysRevD.93.033006.

[45] C L Cowan, F Reines, F B Harrison, H W Kruse, and A D McGuire. Detection of the Free Neutrino: a Confirmation. *Science*, 124(3212):103–4, Jul 1956. ISSN 0036-8075. doi: 10.1126/science.124.3212.103.

[46] Ubaldo Dore, Pier Ferruccio Loverre, and Lucio Ludovici. Measurement of the Weinberg angle in neutrino interactions. *Eur. Phys. J. H*, 41(2):137–155, Jun 2016. ISSN 2102-6459. doi: 10.1140/epjh/e2016-70006-y.

[47] K.A. et al Olive. Review of Particle Physics. *Chinese Phys. C*, 40(10):100001, Oct 2016. ISSN 1674-1137. doi: 10.1088/1674-1137/40/10/100001.

[48] ATLAS Collaboration. Measurement of the $W$-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 78:110, 2018.

[49] ATLAS Collaboration. Measurement of the W±Z boson pair-production cross section in pp collisions at $\sqrt{s} = 13TeV$ with the ATLAS detector. *Phys. Lett. B*, 762:1–22, 2016. doi: http://dx.doi.org/10.1016/j.physletb.2016.08.052.

[50] Measurement of the w+w− production cross section in pp collisions at a centre-of-mass energy of s=13 tev with the atlas experiment. *Physics Letters B*, 773:354 – 374, 2017. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2017.08.047.

[51] ATLAS Collaboration. Measurement of W± and Z-boson production cross sections in pp collisions at with the ATLAS detector. *Phys. Let. B*, 759: 601–521, 2016. doi: 10.1016/j.physletb.2016.06.023.

[52] ATLAS Collaboration. Measurement of the cross section of high transverse momentum $Z \to b\bar{b}$ production in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector. *Phys. Lett. B*, 738:25, 2014. ISSN 0370-2693. doi: 10.1016/j.physletb.2014.09.020.

[53] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider. *Phys. Rev. Lett.*, 100(24):242001, Jun 2008. ISSN 0031-9007. doi: 10.1103/PhysRevLett.100.242001.

[54] ATLAS Collaboration. Combination of searches for $WW$, $WZ$, and $ZZ$ resonances in $pp$ collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Lett. B*, 755(December):285, 2016. ISSN 0370-2693. doi: 10.1016/j.physletb.2016.02.015.

[55] ATLAS Collaboration. Search for pair production of Higgs bosons in the bb¯b. *Phys. Rev. D*, 94(5):052002, 2016. doi: 10.1103/PHYSREVD.94.052002.

[56] CMS Collaboration. Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at √s=13 TeV. *J. High Energy Phys.*, 2017(7), Mar 2017. ISSN 10298479. doi: 10.1007/JHEP07(2017)014.

[57] Search for massive resonances decaying into pairs of boosted W and Z bosons at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-EXO-15-002, CERN, Geneva, 2015.

[58] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, and D. Maître. Next-to-leading order QCD predictions for W + 3 -jet distributions at hadron colliders. *Phys. Rev. D*, 80(7):074036, Oct 2009. ISSN 1550-7998. doi: 10.1103/PhysRevD.80.074036.

[59] Maxime Gabella. The Randall-Sundrum Model. Technical Report June, 2006.

[60] Lisa Randall and Raman Sundrum. Large Mass Hierarchy from a Small Extra Dimension. *Phys. Rev. Lett.*, 83(17):3370–3373, Oct 1999. ISSN 0031-9007. doi: 10.1103/PhysRevLett.83.3370.

[61] CMS Collaboration. Search for high-mass diphoton resonances in proton–proton collisions at 13 TeV and combination with 8 TeV search. *Phys. Lett. B*, 767:147, 2017. doi: 10.1016/j.physletb.2017.01.027.

[62] Kaustubh Agashe, Hooman Davoudiasl, Gilad Perez, and Amarjit Soni. Warped gravitons at the CERN LHC and beyond. *Phys. Rev. D*, 76(3):036006, Aug 2007. ISSN 1550-7998. doi: 10.1103/PhysRevD.76.036006.

[63] ATLAS Collaboration. Search for $WW/WZ$ resonance production in $\ell\nu qq$ final states in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Oct 2017.

[64] Lisa Randall and Raman Sundrum. An Alternative to Compactification. *Phys. Rev. Lett.*, 83(23):4690–4693, Dec 1999. ISSN 0031-9007. doi: 10.1103/PhysRevLett.83.4690.

[65] Duccio Pappadopulo, Andrea Thamm, Riccardo Torre, and Andrea Wulzer. Heavy vector triplets: bridging theory and data. *J. High Energy Phys.*, 2014 (9):60, Sep 2014. ISSN 1029-8479. doi: 10.1007/JHEP09(2014)060.

[66] Michael H. Seymour and Marilyn Marx. Monte Carlo Event Generators. Apr 2013.

[67] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191(1):159–177, 2015. ISSN 00104655. doi: 10.1016/j.cpc.2015.01.024.

[68] Johannes Bellm, Stefan Gieseke, David Grellscheid, Simon Plätzer, Michael Rauch, Christian Reuschle, Peter Richardson, Peter Schichtel, Michael H. Seymour, Andrzej Siódmok, Alexandra Wilcock, Nadine Fischer, Marco A. Harrendorf, Graeme Nail, Andreas Papaefstathiou, and Daniel Rauch. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016. ISSN 14346052. doi: 10.1140/epjc/s10052-016-4018-8.

[69] T Gleisberg, S Höche, F Krauss, M Schönherr, S Schumann, F Siegert, and J Winter. Event generation with SHERPA 1.1. *J. High Energy Phys.*, 2009 (02):007–007, 2009. ISSN 1029-8479. doi: 10.1088/1126-6708/2009/02/007.

[70] Stefan Höche. Introduction to Parton-Shower Event Generators. *Journeys Through Precis. Front. Amplitudes Colliders*, pages 235–295, Nov 2015. doi: 10.1142/9789814678766_0005.

[71] P. Golonka and Z. Was. PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays. *Eur. Phys. J. C*, 45(1):97–107, Jan 2006. ISSN 14346044. doi: 10.1140/epjc/s2005-02396-4.

[72] Gregory Soyez. The SISCone and anti-kt jet algorithms. *Work*, 1:4, Jun 2008.

[73] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet user manual. *Eur. Phys. J. C*, 72(3):1896, 2012. ISSN 1434-6044. doi: 10.1140/epjc/s10052-012-1896-2.

[74] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-$k_T$ jet clustering algorithm. *J. High Energy Phys.*, 2008(04):063–063, Apr 2008. doi: 10.1088/1126-6708/2008/04/063.

[75] Matteo Cacciari, Juan Rojo, Gavin P Salam, and Gregory Soyez. Quantifying the performance of jet definitions for kinematic reconstruction at the LHC. *J. High Energy Phys.*, 2008(12):032–032, Dec 2008. ISSN 1029-8479. doi: 10.1088/1126-6708/2008/12/032.

[76] David Krohn, Jesse Thaler, Lian-Tao Wang, and Lian Tao Wanga. Jet trimming. *J. High Energy Phys.*, 2010(2):84, 2010. ISSN 1029-8479. doi: 10.1007/JHEP02(2010)084.

[77] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches. *Phys. Rev. D - Part. Fields, Gravit. Cosmol.*, 81(9):094023, May 2010. ISSN 15507998. doi: 10.1103/PhysRevD.81.094023.

[78] ATLAS Collaboration. Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s}$=7 TeV using the ATLAS detector. *J. High Energy Phys.*, 2013(9):76, Jun 2013. ISSN 11266708. doi: 10.1007/JHEP09(2013)076.

[79] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. *J. High Energy Phys.*, 2013(9): 29, 2013. ISSN 11266708. doi: 10.1007/JHEP09(2013)029.

[80] Mrinal Dasgupta, Simone Marzani, and Gavin P Salam. QCD calculations for jet substructure. *Nuovo Cim. della Soc. Ital. di Fis. C*, 37(2):131–136, 2014. ISSN 11241896. doi: 10.1393/ncc/i2014-11746-x.

[81] S.D. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Tönnesmann. Jets in hadron–hadron collisions. *Prog. Part. Nucl. Phys.*, 60(2):484–551, 2008. ISSN 01466410. doi: 10.1016/j.ppnp.2007.12.002.

[82] Gavin P. Salam, Lais Schunk, and Gregory Soyez. Dichroic subjettiness ratios to distinguish colour flows in boosted boson tagging. *J. High Energy Phys.*, 2017(3):22, Mar 2017. ISSN 10298479. doi: 10.1007/JHEP03(2017)022.

[83] Jesse Thaler and Ken Van Tilburg. Identifying boosted objects with N-subjettiness. *J. High Energy Phys.*, 2011(3), Nov 2011. ISSN 11266708. doi: 10.1007/JHEP03(2011)015.

[84] Iain W. Stewart, Frank J. Tackmann, and Wouter J. Waalewijn. N jettiness: An inclusive event shape to veto jets. *Phys. Rev. Lett.*, 105(9):92002, Aug 2010. ISSN 00319007. doi: 10.1103/PhysRevLett.105.092002.

[85] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48(7):3160–3166, Oct 1993. ISSN 05562821. doi: 10.1103/PhysRevD.48.3160.

[86] Jesse Thaler and Ken Van Tilburg. Maximizing boosted top identification by minimizing N-subjettiness. *J. High Energy Phys.*, 2012(2):93, 2012. ISSN 1029-8479. doi: 10.1007/JHEP02(2012)093.

[87] Andrew J. Larkoski, Duff Neill, and Jesse Thaler. Jet shapes with the broadening axis. *J. High Energy Phys.*, 2014(4):17, Jan 2014. ISSN 1029-8479. doi: 10.1007/JHEP04(2014)017.

[88] Boosted hadronic top identification at ATLAS for early 13 TeV data. Technical Report ATL-PHYS-PUB-2015-053, CERN, Geneva, Dec 2015.

[89] ATLAS Collaboration. Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at √s = 8 TeV. *Eur. Phys. J. C*, 76(3):154, 2016. ISSN 14346052. doi: 10.1140/epjc/s10052-016-3978-z.

[90] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. Energy correlation functions for jet substructure. *J. High Energy Phys.*, 2013(6), Apr 2013. ISSN 11266708. doi: 10.1007/JHEP06(2013)108.

[91] Andrew J. Larkoski, Ian Moult, and Duff Neill. Power counting to better jet observables. *J. High Energy Phys.*, 2014(12), Sep 2014. ISSN 10298479. doi: 10.1007/JHEP12(2014)009.

[92] Identification of boosted, hadronically-decaying $W$ and $Z$ bosons in $\sqrt{s} = 13$ TeV Monte Carlo Simulations for ATLAS. Technical Report ATL-PHYS-PUB-2015-033, CERN, Geneva, Aug 2015.

[93] ATLAS Collaboration. ATLAS measurements of the properties of jets for boosted particle searches. *Phys. Rev. D*, 86(7):072006, 2012. doi: 10.1103/PhysRevD.86.072006.

[94] Leandro G. Almeida, Seung J. Lee, Gilad Perez, George Sterman, Ilmo Sung, and Joseph Virzi. Substructure of high-pT jets at the LHC. *Phys. Rev. D - Part. Fields, Gravit. Cosmol.*, 79(7), Jul 2009. ISSN 15507998. doi: 10.1103/PhysRevD.79.074017.

[95] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77(7):0–60, Mar 2017. ISSN 14346052. doi: 10.1140/epjc/s10052-017-5004-5.

[96] Expected performance of missing transverse momentum reconstruction for the ATLAS detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-023, CERN, Geneva, Jul 2015.

[97] ATLAS Collaboration. Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 75(1):17, Jan 2015. ISSN 14346052. doi: 10.1140/epjc/s10052-014-3190-y.

[98] Jet mass reconstruction with the ATLAS Detector in early Run 2 data. Technical Report ATLAS-CONF-2016-035, CERN, Geneva, Jul 2016.

[99] ATLAS Collaboration. Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}$ =7 TeV. *Eur. Phys. J. C*, 73(3):2304, Mar 2013. ISSN 14346052. doi: 10.1140/epjc/s10052-013-2304-2.

[100] Jet global sequential corrections with the ATLAS detector in proton-proton collisions at sqrt(s) = 8 TeV. Technical Report ATLAS-CONF-2015-002, CERN, Geneva, Mar 2015.

[101] ATLAS Collaboration. Large-R jet pT and mass scale uncertainties using in-situ track-based measurements, .

[102] Carsten Peterson. Neural networks and high energy physics. *Neural Comput. - NECO*, 5, 1992.

[103] K H. Becks, J Dahm, and F Seidel. Analysing particle jets with artificial neural networks. In Fevzi Belli and Franz Josef Radermacher, editors, *Ind. Eng. Appl. Artif. Intell. Expert Syst. 5th Int. Conf. IEA/AIE-92*, pages 109–112. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992. ISBN 9783540556015. doi: 10.1007/BFb0024961.

[104] B. Denby, C. S. Lindsey, M. Dickson, J. Konigsberg, G. Pauletta, W. Badgett, K. Burkett, and M. Campbell. Performance of the CDF neural network electron isolation trigger at Fermilab. *Nucl. Inst. Methods Phys. Res. A*, 356(2-3): 485–492, Mar 1995. ISSN 01689002. doi: 10.1016/0168-9002(94)01318-7.

[105] Luke de Oliveira, Michael Kagan, Lester Mackey, Benjamin Nachman, and Ariel Schwartzman. Jet-images — deep learning edition. *J. High Energy Phys.*, 2016(7), 2016. ISSN 10298479. doi: 10.1007/JHEP07(2016)069.

[106] Identification of Jets Containing $b$-Hadrons with Recurrent Neural Networks at the ATLAS Experiment. Technical Report ATL-PHYS-PUB-2017-003, CERN, Geneva, Mar 2017.

[107] James Barnard, Edmund Noel Dawe, Matthew J. Dolan, and Nina Rajcic. Parton shower uncertainties in jet substructure analyses with deep neural networks. *Phys. Rev. D*, 95(1), Sep 2017. ISSN 24700029. doi: 10.1103/PhysRevD.95.014018.

[108] Stefano Carrazza. Modeling NNLO jet corrections with neural networks. Apr 2017. ISSN 15095770. doi: 10.5506/APhysPolB.48.947.

[109] Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Matthew D. Schwartz. Pileup Mitigation with Machine Learning (PUMML). *J. High Energy Phys.*, 2017(12), Jul 2017. ISSN 10298479. doi: 10.1007/JHEP12(2017)051.

[110] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(89)90020-8.

[111] Sander Dieleman, Kyle W. Willett, and Joni Dambre. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Mon. Not. R. Astron. Soc.*, 450(2):1441–1459, 2015. ISSN 13652966. doi: 10.1093/mnras/stv632.

[112] Maciej Koch-Janusz and Zohar Ringel. Mutual Information, Neural Networks and the Renormalization Group. 2017.

[113] ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014.

[114] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, Juan Rojo, and Maria Ubiali. Parton distributions with LHC data. *Nucl. Phys. B*, 867(2):244–289, 2013. ISSN 05503213. doi: 10.1016/j.nuclphysb.2012.10.003.

[115] S. Agostinelli, J. Allison, and Et al. GEANT4 - A simulation toolkit. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, 506(3):250–303, 2003. ISSN 01689002. doi: 10.1016/S0168-9002(03)01368-8.

[116] Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector. Technical Report ATLAS-CONF-2015-029, CERN, Geneva, Jul 2015.

[117] Daniel Stump, Joey Huston, Jon Pumplin, Wu-Ki Tung, H L Lai, and Others. Inclusive jet production, parton distributions, and the search for new physics. *Jhep*, 0310(10):46, 2003. doi: 10.1088/1126-6708/2003/10/046.

[118] ATLAS Collaboration. Improved luminosity determination in pp collisions at $\sqrt{s}$ =7 TeV using the ATLAS detector at the LHC. *Eur. Phys. J. C*, 73(8): 2–39, 2013. ISSN 14346052. doi: 10.1140/epjc/s10052-013-2518-3.

[119] Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of $sqrts$ = 7 TeV. Technical Report ATLAS-CONF-2010-038, CERN, Geneva, Jul 2010.

[120] Andrew J. Larkoski, Jesse Thaler, and Wouter J. Waalewijn. Gaining (mutual) information about quark/gluon discrimination. *J. High Energy Phys.*, 2014 (11):129, Nov 2014. ISSN 10298479. doi: 10.1007/JHEP11(2014)129.

[121] Martín Abadi and et al Barham, Paul. TensorFlow: A system for large-scale machine learning. In *12th USENIX Symp. Oper. Syst. Des. Implement. (OSDI 16)*, pages 265–283, GA, 2016. USENIX Association. ISBN 978-1-931971-33-1. doi: 978-1-931971-33-1.

[122] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.*, 15:1929–1958, 2014. ISSN 15337928. doi: 10.1214/12-AOS1000.

[123] Sean C. Smithson, Guang Yang, Warren J. Gross, and Brett H. Meyer. Neural Networks Designing Neural Networks: Multi-Objective Hyper-Parameter Optimization. Nov 2016. ISSN 10923152. doi: 10.1145/2966986.2967058.

[124] R H Hahnloser, R Sarpeshkar, M a Mahowald, R J Douglas, and H S Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789):947–951, Jun 2000. ISSN 0028-0836. doi: 10.1038/35016072.

[125] Daehyun Lee and Kyungsik Myung. Read my lips, login to the virtual world. *2017 IEEE Int. Conf. Consum. Electron. ICCE 2017*, pages 434–435, Dec 2017. ISSN 09252312. doi: 10.1109/ICCE.2017.7889386.

[126] Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS. Technical Report ATLAS-CONF-2015-037, CERN, Geneva, Aug 2015.

[127] Jannicke Pearkes, Wojciech Fedorko, Alison Lister, and Colin Gay. Jet Constituents for Deep Neural Network Based Top Quark Tagging. Apr 2017.

[128] In-situ measurements of the ATLAS large-radius jet response in 13 TeV pp collisions. Technical Report ATLAS-CONF-2017-063, CERN, Geneva, Jul 2017.

[129] Nima Sherafati, Dag Gillberg, Steven Schramm, and Alex Martyniuk. Large-R jet triggers with trimming and mass cuts. Technical Report ATL-COM-DAQ-2017-007, CERN, Geneva, Feb 2017. These plots are intended to be shown at the LHCC poster session on February 22, 2017.

[130] J A Aguilar-Saavedra, Jack Collins, and Rashmish K Mishra. A generic anti-QCD jet tagger. *J. High Energy Phys.*, 2017(11):163, Nov 2017. ISSN 1029-8479. doi: 10.1007/JHEP11(2017)163.

[131] Tao Han, Joseph D. Lykken, and Ren-Jie Zhang. On Kaluza-Klein States from Large Extra Dimensions. *Phys. Rev. D*, 59(10):105006, Mar 1998. ISSN 0556-2821. doi: 10.1103/PhysRevD.59.105006.

[132] Jogesh C. Pati and Abdus Salam. Lepton number as the fourth "color". *Phys. Rev. D*, 10(1):275–289, Jul 1974. ISSN 05562821. doi: 10.1103/PhysRevD.10.275.

[133] Howard Georgi and S. L. Glashow. Unity of all elementary-particle forces. *Phys. Rev. Lett.*, 32(8):438–441, Feb 1974. ISSN 00319007. doi: 10.1103/PhysRevLett.32.438.

[134] John Micklethwait Wooldridge and Adrian. The State of the State. In *Foreign Aff.*, pages 12–15, 2014. ISBN 0887998220111. doi: 10.1038/nrmicro1025.

[135] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rep.*, 516(1-2):1–102, May 2012. ISSN 03701573. doi: 10.1016/j.physrep.2012.02.002.

[136] E. Eichten, I. Hinchliffe, K. Lane, and C. Quigg. Supercollider physics. *Rev. Mod. Phys.*, 56(4):579–707, Oct 1984. ISSN 00346861. doi: 10.1103/RevModPhys.56.579.

[137] J. R. Andersen, O. Antipin, G. Azuelos, L. Del Debbio, E. Del Nobile, S. Di Chiara, T. Hapola, M. Järvinen, P. J. Lowdon, Y. Maravin, I. Masina, M. Nardecchia, C. Pica, and F. Sannino. Discovering technicolor. *Eur. Phys. J. Plus*, 126(9):1–61, 2011. ISSN 21905444. doi: 10.1140/epjp/i2011-11081-1.

[138] D. Barducci, A. Belyaev, S. De Curtis, S. Moretti, and G. M. Pruna. LHC physics of extra gauge bosons in the 4D Composite Higgs Model. *EPJ Web Conf.*, 37(1):265–267, Nov 2013. ISSN 11241896. doi: 10.1393/ncc/i2014-11702-x.

[139] CMS Collaboration. Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at $\sqrt{s} = 8$ TeV. *J. High Energy Phys.*, 2014(8):174, 2014. ISSN 10298479. doi: 10.1007/JHEP08(2014)174.

[140] ATLAS Collaboration. Search for resonant diboson production in the $\ell \ll q\bar{q}$ final state in pp collisions at $\sqrt{s} = 8$ Tev with the ATLAS detector. *Eur. Phys. J. C*, 75(2):69, 2015. ISSN 14346052. doi: 10.1140/epjc/s10052-015-3261-8.

[141] ATLAS Collaboration. Search for production of WW/WZ resonances decaying to a lepton, neutrino and jets in pp collisions at √s = 8 TeV with the ATLAS detector. *Eur. Phys. J. C*, 75(5):209, Mar 2015. ISSN 14346052. doi: 10.1140/epjc/s10052-015-3425-6.

[142] ATLAS Collaboration. Search for a $WZ$ resonance in the fully leptonic channel using $pp$ collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Lett. B*, 737:223, 2014. ISSN 0370-2693. doi: 10.1016/j.physletb.2014.08.039.

[143] Reina Camacho Toro, Mario Campanelli, Joseph Stanford Ennis, Yanyan Gao, Valerio Ippolito, Roland Jansky, Enrique Kajomovitz, Alex Martyniuk, Zachary Alden Meadows, Bill Murray, Attilio Picazio, Steven Schramm, Amal Vaidya, Stephane Willocq, Benjamin Wynne, and Tim Adye. Search for diboson resonances with jets in 36.7 fb $^{-1}$ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATL-COM-PHYS-2016-1490, CERN, Geneva, Oct 2016.

[144] CMS Collaboration. Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s} = 8$ TeV. *J. High Energy Phys.*, 2014(8):173, 2014. ISSN 10298479. doi: 10.1007/JHEP08(2014)173.

[145] Search for resonances with boson-tagged jets in 3.2 fb−1 of p p collisions at √ s = 13 TeV collected with the ATLAS detector. Technical Report ATLAS-CONF-2015-073, CERN, Geneva, Dec 2015.

[146] ATLAS Twiki - CxAODFramework, 2017.

[147] James Catmore, Jack Cranshaw, Thomas Gillam, Eirik Gramstad, Paul Laycock, Nurcan Ozturk, and Graeme Andrew Stewart. A new petabyte-scale data derivation framework for ATLAS. *J. Phys. Conf. Ser.*, 664(7):72007, 2015. ISSN 17426588. doi: 10.1088/1742-6596/664/7/072007.

[148] Zach Marshall. Re-re-defining the Standard QCD Di-Jet Samples: Beginning to like event weights. Technical Report ATL-COM-PHYS-2015-417, CERN, Geneva, May 2015.

[149] Manuel Bähr, Stefan Gieseke, Martyn A. Gigg, David Grellscheid, Keith Hamilton, Oluseyi Latunde-Dada, Simon Plätzer, Peter Richardson, Michael H. Seymour, Alexander Sherstnev, and Bryan R. Webber. Herwig++ physics and manual. *Eur. Phys. J. C*, 58(4):639–707, Dec 2008. ISSN 14346044. doi: 10.1140/epjc/s10052-008-0798-9.

[150] Stefan Gieseke, Christian Röhr, and Andrzej Siódmok. Colour reconnections in Herwig++. *Eur. Phys. J. C*, 72(11):1–18, Nov 2012. ISSN 14346052. doi: 10.1140/epjc/s10052-012-2225-5.

[151] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *J. High Energy Phys.*, 2014(7):79, 2014. ISSN 10298479. doi: 10.1007/JHEP07(2014)079.

[152] Paolo Nason. A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms. *J. High Energy Phys.*, 2004(11):040–040, Nov 2004. ISSN 1029-8479. doi: 10.1088/1126-6708/2004/11/040.

[153] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: The POWHEG method. *J. High Energy Phys.*, 2007(11), Sep 2007. ISSN 11266708. doi: 10.1088/1126-6708/2007/11/070.

[154] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX. *J. High Energy Phys.*, 2010(6), Feb 2010. ISSN 11266708. doi: 10.1007/JHEP06(2010)043.

[155] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at s=13 TeV. *Eur. Phys. J. C*, 76(5): 292, 2016. ISSN 14346052. doi: 10.1140/epjc/s10052-016-4120-y.

[156] Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. Technical Report ATLAS-CONF-2016-024, CERN, Geneva, Jun 2016.

[157] Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2017-004, CERN, Geneva, Apr 2017.

[158] CMS Collaboration. Identification techniques for highly boosted W bosons that decay into hadrons. *J. High Energy Phys.*, 2014(12), Oct 2014. ISSN 10298479. doi: 10.1007/JHEP12(2014)017.

[159] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The catchment area of jets. *J. High Energy Phys.*, 2008(04):5, 2008.

[160] ATLAS Collaboration. Luminoity determination in pp colliion at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC. *Eur. Phys. J. C*, 76(12):653, . doi: 10.1140/epjc/s10052-016-4466-1.

[161] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71(2):1554, 2011. ISSN 14346052. doi: 10.1140/epjc/s10052-011-1554-0.

[162] Alexander L. Read. Presentation of search results: the CL s technique. *J. Phys. G Nucl. Part. Phys.*, 28(10):2693–2704, 2002. ISSN 0954-3899. doi: 10.1088/0954-3899/28/10/313.