

# A Thesis Title

*Author Name*

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy**  
of  
**University College London.**

Department of Something  
University College London

September 6, 2017

I, Author Name, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

# **Abstract**

My research is about stuff.

It begins with a study of some stuff, and then some other stuff and things.

There is a 300-word limit on your abstract.

# Acknowledgements

Acknowledge all the things!

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introductory Material</b>                      | <b>13</b> |
| <b>2</b> | <b>The ATLAS Detector</b>                         | <b>14</b> |
| 2.1      | The Large Hadron Collider . . . . .               | 14        |
| 2.1.1    | LHC running conditions in 2015 and 2016 . . . . . | 14        |
| 2.2      | ATLAS Detector Description . . . . .              | 15        |
| 2.2.1    | ATLAS Co-ordinate System . . . . .                | 16        |
| 2.2.2    | Inner Detector . . . . .                          | 18        |
| 2.2.3    | Calorimeters . . . . .                            | 20        |
| 2.2.4    | Muon Spectrometer . . . . .                       | 23        |
| 2.2.5    | Magnets . . . . .                                 | 24        |
| 2.3      | Trigger . . . . .                                 | 25        |
| <b>3</b> | <b>Object Definition and Reconstruction</b>       | <b>27</b> |
| 3.1      | Jets . . . . .                                    | 27        |
| 3.1.1    | Hadronic topocluster reconstruction . . . . .     | 28        |
| 3.1.2    | Jet Reconstruction . . . . .                      | 29        |

|          |  |           |
|----------|--|-----------|
| 3.1.3    | Jet calibration . . . . .                                | 29        |
| 3.2      | Identification of $b$ -jets . . . . .                    | 30        |
| 3.2.1    | Assigning a Flavour Label . . . . .                      | 30        |
| 3.2.2    | Algorithm descriptions . . . . .                         | 30        |
| 3.2.3    | Performance . . . . .                                    | 34        |
| 3.3      | $b$ Tagging: Validation in Dijet Events . . . . .        | 34        |
| 3.4      | Leptons . . . . .  | 34        |
| 3.4.1    | Electron . . . . .                                       | 34        |
| 3.4.2    | Muon . . . . .   | 34        |
| <b>4</b> | <b>Triggering in the di-<math>b</math>-jet analysis</b>  | <b>35</b> |
| 4.1      | Jet-Triggers . . . . .                                   | 35        |
| 4.1.1    | Level 1 . . . . .  | 36        |
| 4.1.2    | HLT . . . . .  | 36        |
| 4.1.3    | High-mass trigger selection . . . . .                    | 36        |
| 4.2      | $b$ -Jet Triggers . . . . .                              | 37        |
| 4.2.1    | General description . . . . .                            | 37        |
| 4.3      | Efficiency Measurement of the $b$ -Jet Trigger . . . . . | 39        |
| 4.3.1    | Strategy . . . . .                                       | 39        |
| 4.3.2    | Datasets . . . . .                                       | 40        |
| 4.3.3    | Event Selection . . . . .                                | 41        |
| 4.3.4    | The Initial Problem . . . . .                            | 42        |

|  |           |
|--|-----------|
| <i>Contents</i>  | 7         |
| 4.3.5 Investigation . . . . .                                    | 42        |
| 4.3.6 Solution: $b$ -Jet Trigger GRL . . . . .                   | 49        |
| 4.3.7 Efficiency Measurement and Systematic Derivation . . . . . | 51        |
| 4.3.8 Cross-checks . . . . .                                     | 61        |
| 4.4 To Do . . . . .  | 62        |
| <b>5 General Conclusions</b>                                     | <b>63</b> |
| <b>Appendices</b>  | <b>64</b> |
| <b>A An Appendix About Stuff</b>                                 | <b>64</b> |
| <b>B Another Appendix About Things</b>                           | <b>65</b> |
| <b>C Colophon</b>  | <b>66</b> |
| <b>Bibliography</b>  | <b>67</b> |

# List of Figures

|     |   |    |
|-----|---|----|
| 2.1 | Cumulative luminosity versus time delivered to (green) and recorded by ATLAS (yellow) during stable beams for pp collisions at 13 TeV centre-of-mass energy in (a) 2015 and (b) 2016 . . . . .  | 15 |
| 2.2 | A cut-away schematic of the ATLAS detector. . . . .   | 16 |
| 2.3 | A visualisation of the ATLAS detector and the various sub-detectors. The view is taken as a slice in a plane perpendicular to the beam-pipe, showing the radial range from the beam-pipe to the edge of the detector. Overlaid are simplified illustrations of how various types of particles interact with the ATLAS detector; specifically from left to right the particles are an electron, a chargeless hadron (e.g. a neutron), a photon, a muon and a charged hadron (e.g. proton). The sub-detector components are not to scale. . . . . | 17 |
| 2.4 | A cut-away schematic of the ATLAS Inner Detector (ID). . . . .  | 19 |
| 2.5 | A cut-away schematic of the ATLAS calorimeter system. . . . .   | 21 |
| 2.6 | The layout of the ATLAS magnets. . . . .  | 24 |
| 2.7 | A schematic of the ATLAS trigger and data-acquisition system in Run-2, with a focus on the components required for triggering. . . . .  | 25 |
| 3.1 | A schematic illustrating the formation of hadronic jets and the resulting observed energy deposits in the calorimeter. . . . .  | 28 |



|     |  |    |
|-----|--|----|
| 3.2 | A schematic illustrating the algorithm used to form a topocluster. The numbers on the grid represent $ S_{\text{cell}}^{EM} $ and the colours represent the cell label and formed topocluster. . . . .   | 29 |
| 3.3 | A diagram to illustrate the key features of a $b$ -jet that are utilised by the base $b$ -tagging algorithms. . . . .  | 31 |
| 3.4 | A diagram illustrating how three base flavour tagging algorithms are combined in the MV2 algorithm. . . . .  | 33 |
| 4.1 | The expected $b$ -jet efficiency of $b$ -jet triggers with respect to (a) light-jet and (b) $c$ -jet rejection in the case where the $b$ -tagging algorithm used is MV2c20 (Black), IP3D+SV1 (Blue) and for the set-up used in Run-1 (red stars) . . . . .   | 38 |
| 4.2 | The 60% $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for Data (black) and simulation (red) against jet- $p_T$ (a) and jet- $\eta$ (b). The $b$ -trigger aware GRL is not applied and trigger matching is not required. . . . .   | 42 |
| 4.3 | The 60% $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for data from Epoch 1 (black) and simulation (red) against jet- $p_T$ (a) and online beamspot $z$ -position (b). The $b$ -jet trigger aware GRL has not been applied. . . . .   | 45 |
| 4.4 | (a) The 60% $b$ -jet trigger efficiency with respect to an offline 70% operating point tag and (b) the number of offline jets passing 70% operating point tag and matching a HLT trigger jet against vertex class for data from Epoch 1 (black) and simulation (red). The $b$ -jet trigger aware GRL has not been applied. . . . . | 45 |
| 4.5 | The 60% $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for data from epoch 2 (black) and simulation (red) against jet- $p_T$ (a), jet- $\eta$ (b) and online beamspot $z$ -position (c). . . . .   | 47 |

- 4.6  $b$ -perf efficiency,  $\epsilon_{bPerf}$ , for data from Epoch 2 (black) and simulation (red) against leading-jet  $p_T$  (a), online beamspot  $z$ -position (b) and leading jet- $\eta$  (c). . . . . 47
- 4.7 The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for data from Epoch 3 (black) and simulation (red) against (a) jet- $p_T$ , (b) jet- $\eta$ , (c) online beamspot  $z$ -position and (d) vertex class. . . . . 48
- 4.8 The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for data from Region 1 (black) and simulation (red) against jet- $p_T$  (a) and jet- $\eta$  (b). The  $b$ -jet trigger aware GRL has been applied. . . . . 50
- 4.9  $b$ -perf efficiency,  $\epsilon_{bPerf}$ , for data from Region 2 (black) and simulation (red) against leading (a) jet- $p_T$  and (b) jet- $\eta$ . The  $b$ -jet trigger aware GRL has been applied. . . . . 50
- 4.10 The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for the full 2016 data-set (black) and simulation (red) against jet- $p_T$  (a), jet- $\eta$  (b), online beamspot  $z$ -position (c) and vertex class (d). . . . . 51
- 4.11  $b$ -perf efficiency,  $\epsilon_{bPerf}$ , for the full 2016 data-set (black) and simulation (red) against (a) leading jet- $p_T$  and (b) jet- $\eta$ . The  $b$ -jet trigger aware GRL has been applied. . . . . 51
- 4.12 A comparison of offline jets tagged at the 70% operating point for inclusive jets (red) and truth-matched  $b$ -jets (blue) against jet- $p_T$  (a) and jet- $\eta$  (b) in a simulated  $t\bar{t}$  sample. . . . . 53
- 4.13 The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for inclusive jets (black) compared to truth matched  $b$ -jets and non  $b$ -jets (a) and the case where non  $b$ -jet content has been doubled (b) for a simulated  $t\bar{t}$  sample. The lower panel in both plots show the ratio to the inclusive efficiency. . . . . 53

- 4.14 The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for nominal inclusive case (black) compared to varied inclusive case (red) and just non  $b$ -jets (blue) in the case where non  $b$ -jet efficiency has been halved (a) and doubled (b) for a simulated  $t\bar{t}$  sample. The lower panel in both plots show the ratio of the varied inclusive efficiency to the nominal inclusive efficiency. . . . . 54
- 4.15 A figure to demonstrate the high- $p_T$  extrapolation procedure for the 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag. Data (black) is compared against simulation (red) after various corrections have been applied as a function of jet- $p_T$ . Panel (a) shows the the flat normalisation fit uncorrected simulation, panel (b) shows the linear correction fit to normalised simulation, panel (c) shows the linear correction fit errors to the corrected simulation and panel (d) shows the quadratic fit to the corrected simulation. . . . . 56
- 4.16 The measured 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag as measured in data as a function of offline jet- $p_T$ . The central values are shown in black with the statistical error and the green bands represent the total error including systematic errors. . . . . 58
- 4.17 Data/simulation scale factors for the 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag as a function of offline jet- $p_T$ . The central values are shown in black with the statistical error and the green bands represent the total error including systematic errors. . . . . 58
- 4.18 The measured 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag as measured in data (black) and from the corrected simulation (red) as a function of offline jet- $p_T$ . In the ratio plot on the lower panel the extrapolation errors is represented by the red band, whilst the total error is overlaid in green. . . . . 59
- 4.19 The measured  $\epsilon_{bPerf}$  as measured in data as a function of offline leading jet- $\eta$ . The central values are shown in black with the statistical error and the green bands represent the total error including systematic errors. . . . . 61

# List of Tables

|     |   |    |
|-----|---|----|
| 3.1 | The Mv2c10 $b$ -tagging algorithm operating points; with the corresponding $b$ -jet efficiency, $c$ -jet rejection, light-jet rejection and $\tau$ rejection. . . . .   | 33 |
| 4.1 | A table describing the effect of not finding a valid xPrmVtx primary vertex on different epochs of data. . . . .  | 43 |
| 4.2 | A table showing the systematic assigned for the high- $p_T$ extrapolation. . . .  | 55 |
| 4.3 | A table showing the jet-level Data/simulation scale factor (SF) as a function of jet- $p_T$ with total error and the contributions of the different systematics considered; specifically statistical, high- $p_T$ extrapolation, non- $b$ -jet purity and non- $b$ -jet trigger efficiency. . . . . | 57 |
| 4.4 | A table showing the event-level Data/MC scale factor (SF) as a function of leading jet- $\eta$ with total error and the contributions of the different systematics considered. . . . .  | 61 |

## Chapter 1

# Introductory Material

Some stuff about things.[?] Some more things.

Inline citation:

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

## Chapter 2

# The ATLAS Detector

## 2.1 The Large Hadron Collider

High-energy particle colliders have been an essential tool in high-energy physics research for over 50 years, with a rich history of discovering new particles as each generation of collider pushes the energy frontier; including the discovery of the Z and W bosons using the Super Proton Synchrotron at CERN in 1983 [1, 2, 3, 4] and the discovery of the top-quark at the Tevatron in 1995 [5, 6].

The Large Hadron Collider (LHC) is the highest energy collider ever built, operated by the *Conseil Européen pour la Recherche Nucléaire (CERN)*. Lying in a tunnel 100 m beneath the Swiss/French border near Geneva, the LHC is a 27 km circumference ring of superconducting magnets and accelerating structures, which accelerate beams of protons to a maximum energy of 6.5 TeV. These proton beams are collided in four different locations on the LHC ring and around each collision point a different detector is constructed to observe these collisions; one such of these detectors is ATLAS.

### 2.1.1 LHC running conditions in 2015 and 2016

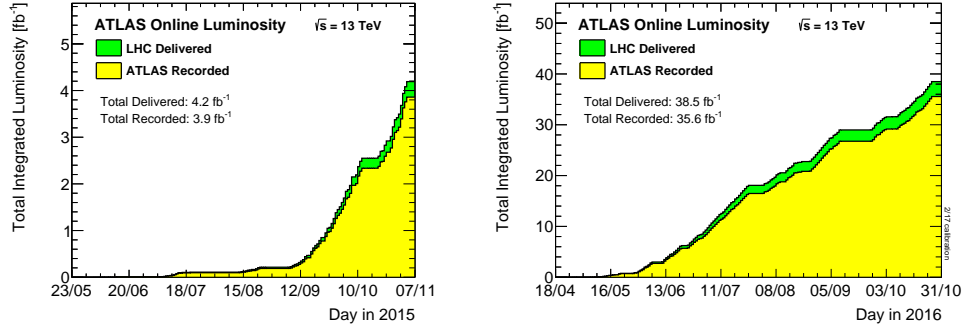
Since May 2015 the LHC has been colliding bunches of protons at a center-of-mass energy of 13 TeV, the highest energy collisions ever obtained by a particle collider <sup>1</sup>. In 2015 and 2016 the LHC produced pp collisions with a bunch spacing of 25 ns<sup>2</sup> and an average number

---

<sup>1</sup>The period of data-taking from 2015 is known as Run-2.

<sup>2</sup>A small amount of data in 2015 was collected with a bunch spacing of 50 ns

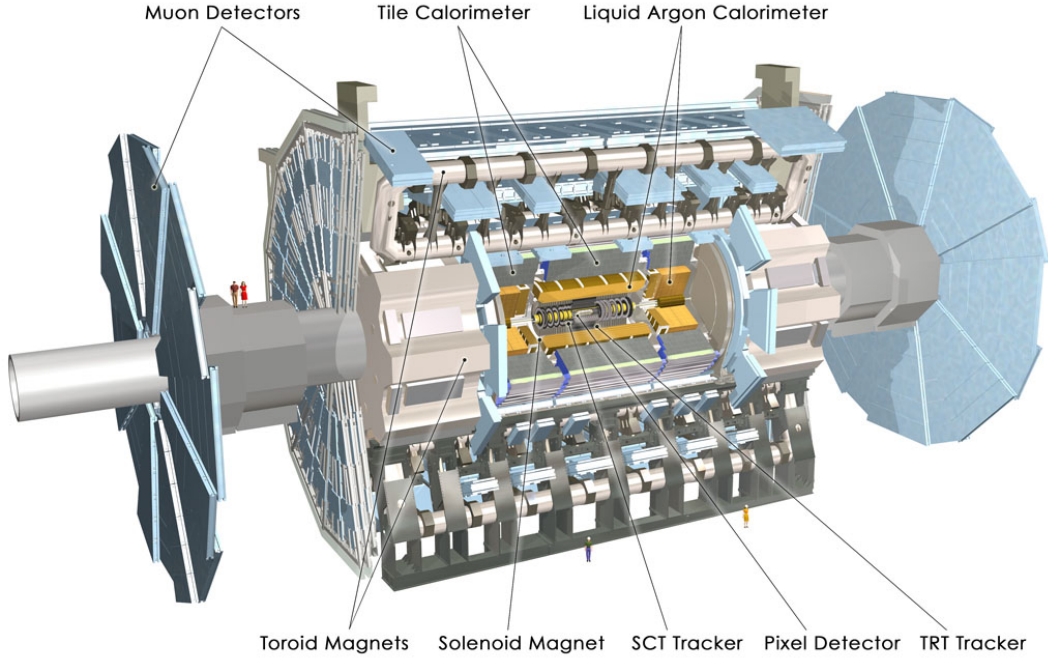
of collisions per bunch-crossing ( $\langle \mu \rangle$ ) of 23.7. Figure 2.1 shows the total luminosity delivered by the LHC and recorded by ATLAS against date in 2015 and 2016, showing that a luminosity of  $39.5 \text{ fb}^{-1}$  was recorded by ATLAS in 2015 and 2016 combined [7].



**Figure 2.1:** Cumulative luminosity versus time delivered to (green) and recorded by ATLAS (yellow) during stable beams for pp collisions at 13 TeV centre-of-mass energy in (a) 2015 and (b) 2016 [7].

## 2.2 ATLAS Detector Description

The ATLAS (**A** **T**oroidal **L**arge **H**adron Collider **A**pparatu**S**) detector design, construction and performance has been described in detail previously [8, 9, 10], so what follows in this chapter is a general description of the detector with a focus on the needs of the analysis that is being presented. The ATLAS detector is effectively a large closed cylindrical detector, made up of four key components which sit in concentric rings around the interaction point, where the proton bunches collide. These components are the inner detector, calorimeters, muon spectrometer and the magnets; each of which are described in further detail below. This design is used as each sub-detector measures different quantities and interacts differently to the various range of particles that ATLAS is required to observe, meaning the ATLAS detector is able to identify and measure the key properties of particles that pass through its volume. Figure 2.2 shows a cut-away schematic of the detector and Figure 2.3 shows a slice of the detector in the plane perpendicular to the beam-pipe, overlaid are simplified illustrations how the detector can respond to a range of particles [11].



**Figure 2.2:** A cut-away schematic of the ATLAS detector [8].

### 2.2.1 ATLAS Co-ordinate System

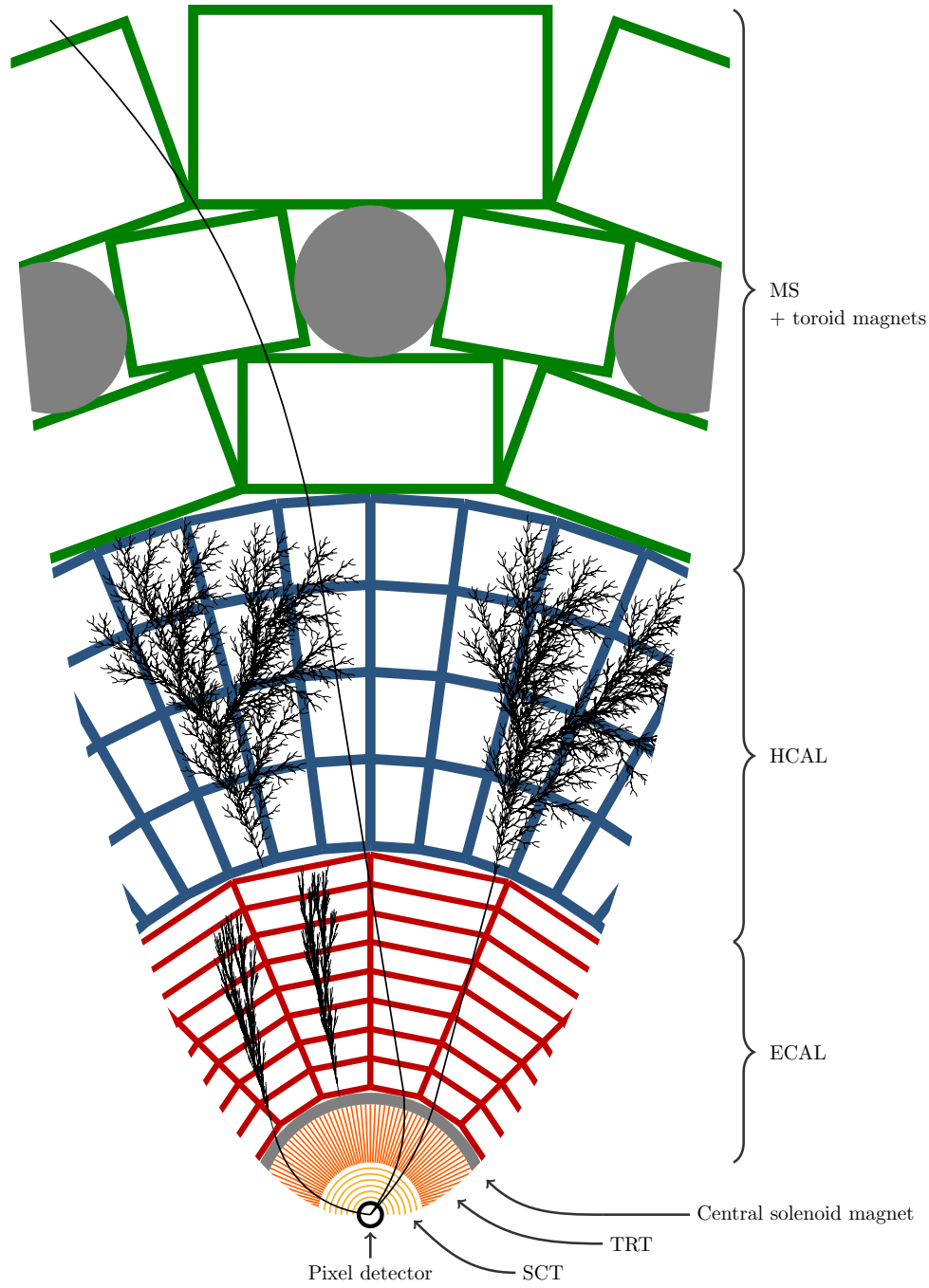
Firstly, to describe the detail of the ATLAS detector there must be a description of the co-ordinate system that is used. ATLAS uses a right-handed coordinate system, in which the origin lies at the interaction point. The  $x$ -axis points to the centre of the LHC ring parallel to the surface of the earth, the  $y$ -axis points towards the surface of the earth and the  $z$ -axis runs along the beam-pipe, pointing anti-clockwise along the LHC beam-pipe. The azimuthal angle,  $\phi$ , is defined right-handedly around the  $z$ -axis starting at the  $x$ -axis.

The polar angle,  $\theta$ , is defined as the angle measured from the  $z$ -axis, such that along the  $z$ -axis corresponds to  $\theta = 0$  and anti-aligned with the  $z$ -axis corresponds to  $\theta = \pi$ . However, to define the angular direction with respect to the  $z$ -axis the ATLAS co-ordinate system uses pseudo-rapidity,  $\eta$ , instead of using  $\theta$ , for reasons that will be outlined below.  $\eta$  is defined as a function of  $\theta$ :

$$\eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right] \quad (2.1)$$

Thus,  $\eta = 0$  corresponds to a particle travelling perpendicular to the beam-pipe, where a positive value of  $\eta$  corresponds to a particle travelling with a tilt towards the  $z$ -axis. The quantity is called pseudo-rapidity as in the massless limit ( $\lim_{E \rightarrow |\vec{p}|}$ ) it can be shown that  $\eta$





**Figure 2.3:** A visualisation of the ATLAS detector and the various sub-detectors. The view is taken as a slice in a plane perpendicular to the beam-pipe, showing the radial range from the beam-pipe to the edge of the detector. Overlaid are simplified illustrations of how various types of particles interact with the ATLAS detector; specifically from left to right the particles are an electron, a chargeless hadron (e.g. a neutron), a photon, a muon and a charged hadron (e.g. proton). The sub-detector components are not to scale [11].

converges to rapidity,  $y$ , where rapidity is defined as,

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right) \quad (2.2)$$

A key property of rapidity is that the differences in rapidity,  $\Delta y$ , are invariant against Lorentz boosts along the  $z$ -axis. Thus,  $\eta$  is the final variable chosen in the ATLAS co-ordinate system due to the relation of  $\eta$  with both  $\theta$  and  $y$  and the above mentioned property of  $\Delta y$ . One final quantity commonly used within ATLAS is the variable  $\Delta R$ , which is defined as

$$\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} \quad (2.3)$$

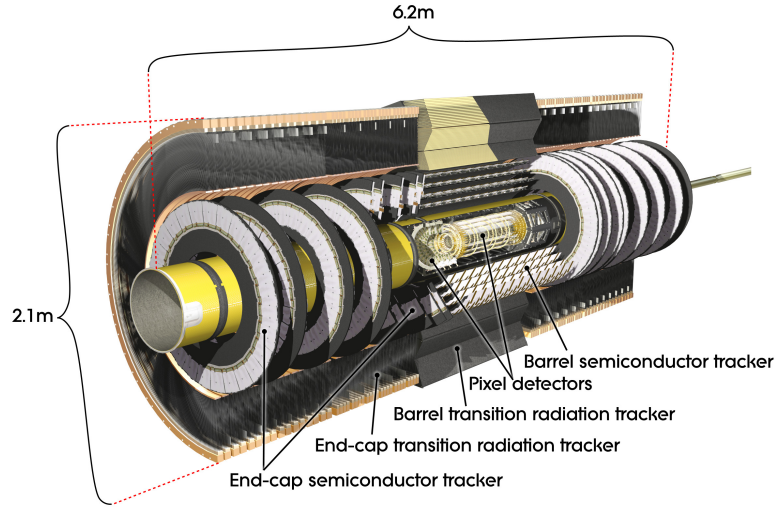
$\Delta R$  represents an angular separation between two vectors within the ATLAS co-ordinate system.

Now that we have discussed the ATLAS co-ordinate system, we can provide a description of the components of the ATLAS detector.

### 2.2.2 Inner Detector

The Inner Detector (ID), the innermost sub-detector on ATLAS, measures the trajectory of charged particles passing through the detector. The ID is constructed from many concentric layers of detector, and as a charged particle passes through the detector each of the layers provides a position measurement, known as a hit. Then using the hits from the many layers the trajectory of the particle can be determined; the measured trajectory is known as a track. The ID is immersed in a 2 T magnetic field which bends the particle's trajectories; from the sign and magnitude of the track's curvature the charge and momentum of the particle can be inferred. The ID is made of three main component parts; the pixel detector, the Semi-Conductor Tracker (SCT) and the Transition Radiation Tracker (TRT), as visualised in Figure 2.4. The ID consists of the barrel, which are made up of cylinders surrounding the beam-pipe to cover low absolute values of  $\eta$ , and the end-caps, which lie perpendicular to beam-pipe on either end of the barrel to cover large values of absolute  $\eta$ : here the description focuses on the barrel as this covers the  $\eta$  range considered by the analysis.

The innermost component of the ID is the silicon pixel detector; in the barrel this detector consists of 4 high-granularity layers of silicon based pixel modules surrounding



**Figure 2.4:** A cut-away schematic of the ATLAS Inner Detector (ID) [8].

the beam pipe, covering a range of  $-2.5 < \eta < 2.5$  and a radial distance of 33 mm to 122.5 mm [12, 13]. The high-granularity of the pixel layers, allows for high precision measurements, with an intrinsic resolution of approximately resolution of  $\sim 10 \mu\text{m}$  in  $R - \phi$  plane and  $\sim 115 \mu\text{m}$  in the  $z$ -direction.

Moving radial outwards the next component of the ID is the Semi-Conductor Tracker; which, in the barrel, comprises of 4 cylindrical layers of silicon micro-strips covering a range of  $-2.5 < \eta < 2.5$  and a radial distance of 299 mm to 514 mm. The SCT has an intrinsic resolution of  $\sim 17 \mu\text{m}$  in  $R - \phi$  plane and  $\sim 580 \mu\text{m}$  in the  $z$ -direction.

The outermost component of the ID is the Transition Radiation Tracker (TRT) constructed of many 4 mm radius tubes filled with xenon. As a charged particle passes through the gas, it will cause ionisation allowing a measurement of its position using drift-time. In the barrel, each tube provides a measurement in the  $R - \phi$  plane with an intrinsic resolution of  $130 \mu\text{m}$  and the TRT will typically provide 36 hits per track. In addition to a position measurement, due to the choice of the material between the tubes, a particle passing through the detector will radiate photons at an intensity inversely correlated to the mass of that particle, providing additional information for particle identification.

The trajectory, momentum and charge measurements provided by the Inner Detector are essential for particle identification in ATLAS. In particular, the high precision measurements close to the beam-line allow for vertex reconstruction, which is essential for

identification of tracks coming from B or C hadrons, and hence the identification of  $b$ -jets. This process, known as  $b$ -tagging, is discussed further in Section 3.2(*object definition and selection*) and is important within the context of this analysis.

### 2.2.3 Calorimeters

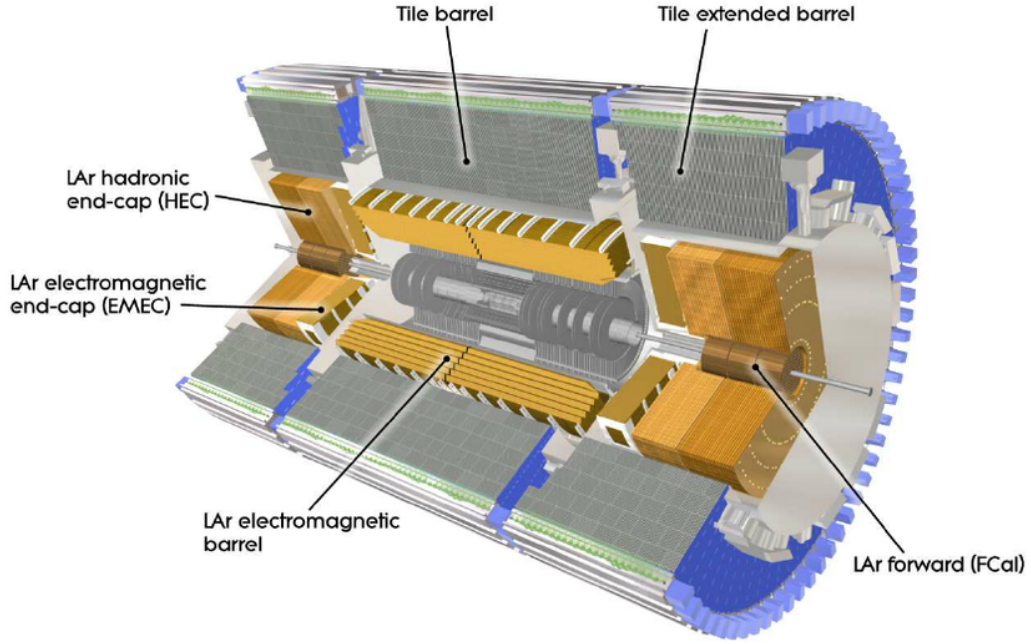
The ATLAS calorimeter, located on the outside of the magnet solenoid surrounding the ID, is designed to provide an energy measurement of the traversing particles. Accurate energy measurements are essential for a good resolution of the mediator mass reconstructed from its decay products, which is important within the context of the analysis being presented in this thesis.

The calorimeter at ATLAS is made up of two different systems that are built in concentric rings; the inner-most is the Electromagnetic Calorimeter system (ECAL), which is used to measure electromagnetic objects such as photons and electrons. Outside of that is the Hadronic Calorimeter system (HCAL), designed to provide an energy measurement of hadronic material. The HCAL is built from the Tile and Hadronic Endcap calorimeters. Both the ECAL and HCAL have barrel and end-cap components to make energy measurements at a large range of  $\eta$  values. Figure 2.5 shows a cut-away of the ATLAS calorimeter.

Below I provide a more detailed description of the calorimeter components; however, the principle behind each detector is common so is described first. The calorimeters at ATLAS are sampling calorimeters, which means they consist of alternating layers of absorber and active material. The role of the absorber layer is to force the particle, whose energy we want to measure, to emit secondary particles. These secondary particles will again emit further particles and so on meaning a “particle cascade” is formed. The role of the active material layer is to measure the energy of the many resulting particles from the cascade, known as the cascade particles. The ATLAS detector is built such that the initial particle will cascade within the volume of the calorimeter system and then, from a measurement of the energy of all the cascade particles, the energy of the initial particle can be inferred.

#### 2.2.3.1 Electromagnetic Calorimeter (ECAL)

For the electromagnetic interaction, at energies  $\sim \geq 1$  GeV the particle cascade process is mainly caused by two processes; bremsstrahlung, ( $e^{+/-} \rightarrow e^{+/-} + \gamma$ ) and pair production



**Figure 2.5:** A cut-away schematic of the ATLAS calorimeter system [8].

( $\gamma \rightarrow e^+ + e^-$ ). The electromagnetic calorimeter at ATLAS is known as the Liquid Argon (LAr) calorimeter. The absorber material used in the LAr calorimeter is lead, due to its large density of charged particles (high  $Z$ ) which increases the rate of the cascade processes. The active material is liquid argon; when a cascade particle passes through the liquid argon it causes ionisation, and the released electrons are captured using an electric field. The number of released electrons is proportional to the energy of the cascade particle, meaning that the energy of the cascade particle can be measured.

As discussed above the LAr is split up into two sections; the barrel section covers a region of  $|\eta| < 1.475$  and two end-cap components cover  $1.375 < |\eta| < 3.2$ . The depth of an electromagnetic calorimeter is often expressed in terms of the radiation length,  $X_0$ , which is the distance that an electron's energy reduces by a factor of  $e^{-1}$  through bremsstrahlung, or  $7/9$  of the mean free path for a photon to pair produce electrons. It is worth noting that this quantity is strongly material dependant; a high- $Z$  material, such as lead, has a shorter  $X_0$ . The LAr calorimeter has a depth of  $> 22 X_0$  in the barrel and  $> 24 X_0$  in the end-caps, meaning that almost all of the particle shower from a high-energy photon or electron can be contained within electromagnetic calorimeter.

### 2.2.3.2 Hadronic Calorimeter (HCAL)

If a particle can also interact through strong interactions, such as the components of a hadronic jet, then the particle cascade is a more complicated process. A hadronic cascade process is dominated by processes such as ionisation, nuclear spallation and neutron generation [14, 15]. For a chargeless hadron, for example a neutron, strong processes, such as spallation, are the only processes that contribute to its cascade. During these hadronic cascade processes many  $\pi_0$  mesons are made, which can decay to a pair of photons and thus form electromagnetic cascade as described above.

For hadronic interactions, the size of detector is measured by the interaction length,  $\lambda$ , defined as the distance required to reduce the number of relativistic hadrons by  $e^{-1}$ . This means that by the end of the LAr calorimeter there is  $2.3 \lambda$  of active material in the barrel, so the full hadronic shower cannot be captured by the LAr calorimeter alone. For a full measurement of the hadronic energy, the Hadronic Calorimeter system (HCAL) is required.

The Tile Calorimeter is constructed from absorber layers of steel and active material layers of scintillating tiles, and has a depth of  $7.4 \lambda$ , meaning the majority of the hadronic shower can be captured by either the LAr calorimeter or the Tile calorimeter. The Tile Calorimeter is split up into the barrel and the extended barrel components; the barrel covers the region  $|\eta| < 1.0$  and the extended barrel covers the region  $0.8 < |\eta| < 1.7$ .

To cover the more forward regions there are two more calorimeter detectors. The Hadronic Endcap Calorimeter (HEC) is housed in two large wheels at either end of the ATLAS detector and covers a region of  $1.5 < \eta < 3.2$ . The HEC is a sampling calorimeter built using copper as the absorber layers and liquid argon as the active material and has a depth of  $\sim 12 \lambda$ . In addition the Forward Calorimeter (FCAL) covers the very forward region of  $3.1 < \eta < 4.9$ , which is outside the range considered within this analysis. It is constructed from absorber layers of copper (for EM interactions) and tungsten (for hadronic interactions) with liquid argon for the active material layers.

Another important point about the ATLAS calorimeter is that it is non-compensating calorimeter; that is to say that the response of the detector to an electromagnetic particle (such as an electron) is larger than the response of a hadronic particle (for example a pion). The reason for this is some energy is lost in hadronic cascade process; mainly due to the

energy required to release nucleons from calorimeter nuclei during spallation, but also from the recoil energy given to the calorimeter nuclei and neutrinos created during strong processes that can escape the calorimeter [16, 17]. To account for the fact that the ATLAS calorimeter is non-compensation, calorimeters are calibrated to the EM-scale and then a jet energy scale correction is applied, this process is described further in Section 3.1(*object definition and selection*).

#### 2.2.4 Muon Spectrometer

The only standard model particle visible to ATLAS which can pass through the calorimeter is the muon; hence to identify and obtain the momentum of muons an additional detector, the Muon Spectrometer (MS), is used. The MS is a detector which surrounds the hadronic calorimeter, measuring the momentum of muons by observing the curvature of their trajectories in magnetic fields. Trajectories are determined using muon position measurements from multiple layers of detectors, analogous to what has been described for the inner detector.

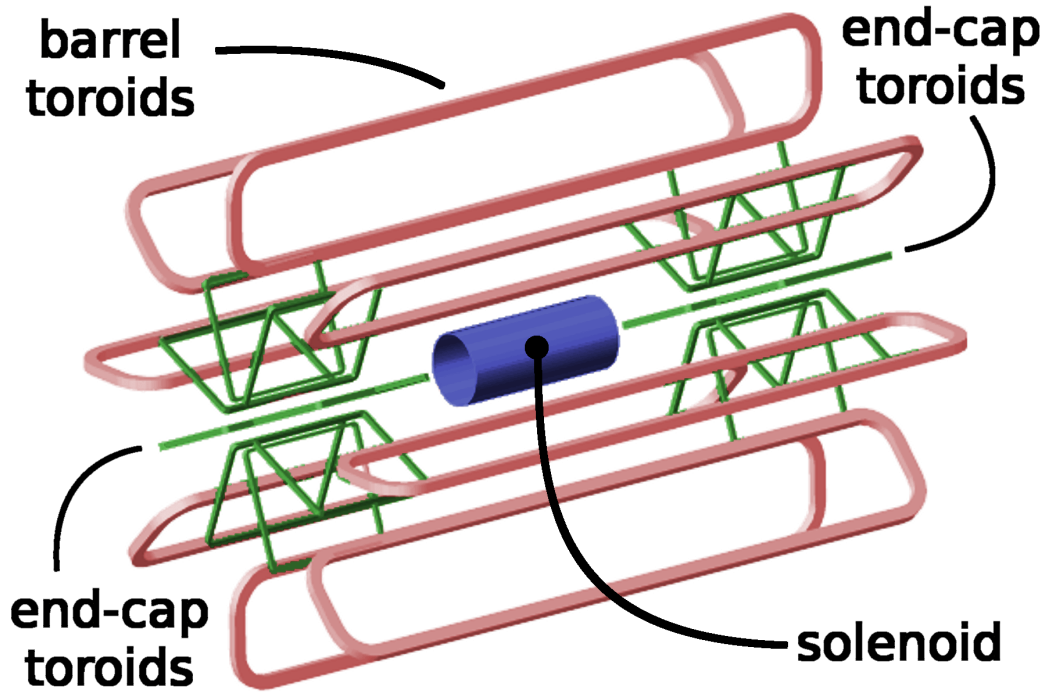
In the barrel region ( $|\eta| < 1.4$ ) the large barrel toroid provides the magnetic field, in the end-cap region ( $1.6 < |\eta| < 2.7$ ) the two smaller end-cap magnets provide the magnetic field and finally in the transition region ( $1.4 < |\eta| < 1.6$ ) both sets of magnets contribute to the magnetic field. A further description of the magnets used in ATLAS is found in the next section.

Muon chambers are the detectors tasked with providing the muon position measurements required to reconstruct muon tracks. The muon chambers come in two types; trigger and precision. The trigger muon chambers provide a quick position measurement in 3-dimensions which can be used to identify muons tracks in the trigger. The trigger muon chambers cover a range  $|\eta| < 2.0$ ; consisting of Resistive Plate Chambers (RPCs) in the barrel and Thin Gap Chambers (TGCs) in the end-cap regions. The precision muon chambers provide a precise measurement of the muon position co-ordinates in the  $R - z$  plane, the plane in which track curvature occurs in the muon spectrometer, allowing for precise measurements of the muon track- $p_T$ . In the barrel region, precision muon chambers are arranged in three concentric cylindrical layers of chambers formed around the beam-pipe, whilst in the transition and end-cap regions there are three layers of chambers either side

of the barrel lying in disks perpendicular to the beam-pipe. In the region  $|\eta| < 2.0$ , the precision muon chambers are made from Monitored Drift Tubes (MDTs), whilst at large pseudo-rapidities ( $2.0 < |\eta| < 2.7$ ), Cathode Strip Chambers (CSCs) are used.

There is an additional use of the muon spectrometer that relates to high-energy jets. Whilst for most jets their shower is fully contained within the calorimeter there are some jets, particularly at high- $p_T$ , where a non-negligible fraction of energy from the shower goes beyond the calorimeter. This effect, known as ‘punch-through’, is accounted for using energy deposits in the muon spectrometer.

### 2.2.5 Magnets



**Figure 2.6:** The layout of the ATLAS magnets [18].

In ATLAS magnetic fields are important for obtaining the momentum and charge of particles from their observed trajectories in the ID and Muon Spectrometer. ATLAS is made up of four large superconducting magnets; the inner solenoid which surrounds the inner detector and provides a 2 T magnetic field within the ID. The barrel toroid magnet provides a magnetic field of up to 2.5 T in the central regions of the muon spectrometer and the two end-cap toroid magnets which produce a magnetic field of up to 3.5 T in the

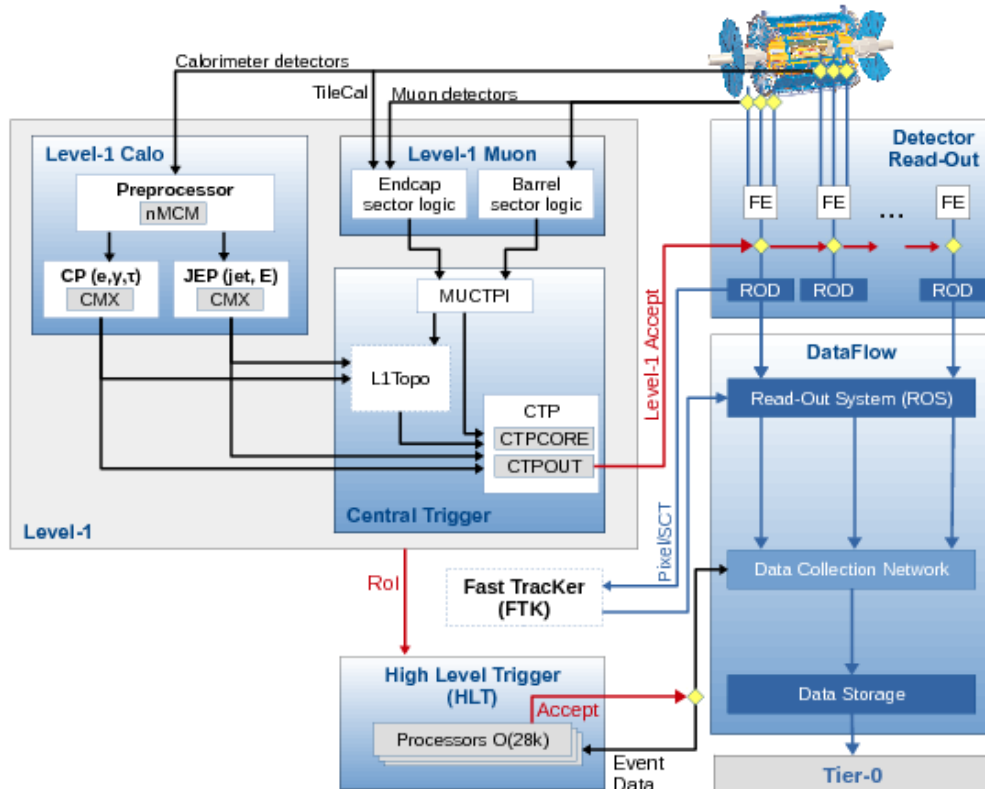


forward regions of the MS. Figure 2.6 shows the layout of the magnets in ATLAS [18].

## 2.3 Trigger

In 2015 and 2016, the LHC has been colliding proton beams with a spacing of 25 ns, meaning that the ATLAS experiment has been taking data at a rate of 40 MHz. However, due to the large computing resources required to process and store each event, it is not possible to record all this data for use in an analysis. To resolve this problem, the ATLAS experiment uses a trigger system to reduce the event rate by selecting the events of interest that contain high- $p_T$  physics objects, which indicate that a hard scatter has occurred in that event.

The ATLAS trigger-system has two levels; the first level trigger (L1) and the higher level trigger (HLT) [19]. Figure 2.7 shows a schematic outlining the trigger used in Run-2 [20].



**Figure 2.7:** A schematic of the ATLAS trigger and data-acquisition system in Run-2, with a focus on the components required for triggering [19].

The first level trigger (L1) is hardware based which reduces the rate from 40 MHz to 100 kHz within a time window of  $2.2\ \mu\text{s}$ . The L1 trigger uses custom electronics to rapidly process information directly from the calorimeter and the muon spectrometer, searching for high- $p_T$  muon tracks and large calorimeter deposits. The information is then passed to the central trigger which uses a set of pre-defined conditions to decide if a L1 trigger accept is given and thus events are passed on to the next step of triggering. At the same time Regions of Interests (ROIs) are constructed around the objects that have fired the L1 trigger, which are passed on to the HLT.

The next step is the HLT, a software based trigger, which further reduces the event rate to 1 kHz within a time window of 0.2 s. The HLT uses the information from the full detector to perform a more complete reconstruction of the physics objects within the event, the most time consuming reconstruction algorithms only being run only within the ROIs taken from L1. The more complex event analysis allowed within the software-based trigger includes track reconstruction and therefore allows for  $b$ -jet identification. If the content of the event reconstruction passes a pre-set criteria, a HLT accept is issued meaning that the events are passed on for processing and storage.

A further description of triggers used in the analysis, with a particular focus on the  $b$ -jet trigger performance can be found in ??(*b-jet trigger chapter*).

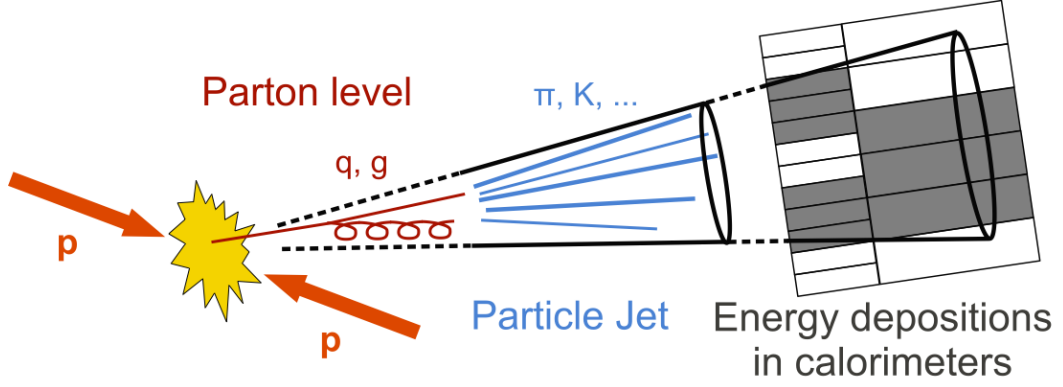
## Chapter 3

# Object Definition and Reconstruction

### 3.1 Jets

If a collision results in a free quark or gluon in its final state then a stream of high-energy hadrons is known as a hadronic jet is formed, the formation of a hadronic jet has been described in described in in Section *[QCD theory description]* *Not written yet* **LM fix/** To summarise the process; firstly the free quark/gluon will radiate additional gluons and quarks in a process known as the parton shower, then these gluons and quarks will then undergo hadronisation to form hadrons which are the constituents of the hadronic jet. The components of the hadronic jet deposit energy in the cells of the ATLAS calorimeter, through the processes described in Section 2.2.3, such that the ATLAS calorimeter has an energy measurement of the components of the hadronic jet, within the spatial limits from the granularity detector. This process of hadronic jet formation and energy deposition in the calorimeter is illustrated in Figure ??.

This section contains a description of the procedure utilised by ATLAS to go from energy deposits in calorimeter cells to well defined and calibrated hadronic jets. This procedure can be split up into three separate steps that are described in the following sections; firstly topoclusters are formed as described in Section 3.1.1, then jets are formed from topoclusters using reclustering algorithms described in Section 3.1.2 and finally Section 3.1.2 describes how the jets are calibrated and the relevant jet energy uncertainties are derived.



**Figure 3.1:** A schematic illustrating the formation of hadronic jets and the resulting observed energy deposits in the calorimeter [21].

### 3.1.1 Hadronic topocluster reconstruction

The first step of jet building at ATLAS builds 3D clusters, known as topoclusters, from neighbouring groups of energy deposits in cells [22]. The calorimeter cells can be from either the EM or hadronic calorimeter systems, which have a dimensions given in Tables **LM fix, need to make tables, then reference here**. The algorithm employed makes use of the variable cell signal significance defined as,

$$S_{\text{cell}}^{EM} = \frac{E_{\text{cell}}^{EM}}{\sigma_{\text{noise,cell}}^{EM}} \quad (3.1)$$

where  $E_{\text{cell}}^{EM}$  is the energy deposited in a cell at the EM-scale **Laurie fix, check EM scale is defined possibly in detector** and  $\sigma_{\text{noise,cell}}^{EM}$  is the uncertainty due to noise in that cell. The sources of noise in a calorimeter cell are described in Section **LM fix, need to have a noise section somewhere....** A large value of  $S_{\text{cell}}^{EM}$  ( $> 1$ ) indicates that the energy deposit is likely due to a particle depositing energy in the calorimeter rather than noise within the calorimeter.

Using the value of  $S_{\text{cell}}^{EM}$ , each calorimeter cell is labelled as follows

- If  $|S_{\text{cell}}^{EM}| > 4$ : the cell is labelled a **seed** cell.
- If  $|S_{\text{cell}}^{EM}| > 2$ : the cell is labelled a **growth** cell.
- If  $|S_{\text{cell}}^{EM}| > 0$ : the cell is labelled a **boundry** cell.

Then the algorithms builds clusters as in the following steps,



## 3.2 Identification of $b$ -jets

Hadronic jets, described in Section 3.1, can be further categorised into three separate categories based on the flavour of the constituent quarks.  $b$ -jets are defined as jets containing one or more  $b$ -hadrons,  $c$ -jets are defined as jets containing one or more  $c$ -hadrons but no  $b$ -hadrons and finally light-flavoured jets comprise of only light hadrons ( $u$ ,  $d$  and  $s$  quarks). A description of how this definition is practically used in simulation is given in Section 3.2.1.

The identification  $b$ -jets, known as  $b$ -tagging, has been an essential tool in a range of ATLAS collaboration results; for example analyses studying the  $t\bar{t}$  final state [?] <sup>1</sup> and the first direct evidence of the Higgs boson coupling to the quark-sector [?]. In the same sense, identification of  $b$ -jets is an essential part of the analysis being presented here; by selecting  $b$ -jets we increase our sensitivity to BSM models that decay to 1 or 2  $b$ -jets in their final state. **Laurie Fix, link to where I explain why this is good, maybe Intro.**

The process of  $b$ -tagging at ATLAS in Run-2 has been previously described in great detail [?, ?], so what follows is a summary of the key features of the process.

### 3.2.1 Assigning a Flavour Label

In simulation, the particle-level truth information is known, and hence a truth flavour label of a jet can be defined. Flavour is assigned to jets by matching truth-level heavy-hadrons with  $p_T > 5$  GeV and  $\Delta R < 0.3$  between the hadron and the jet. If a  $b$ -hadron is matched to a jet, the jet is then declared a  $b$ -jet; this process is then repeated for  $c$ -hadrons and then  $\tau$  leptons. If no match between  $b$ ,  $c$  or  $\tau$  is achieved then the jet is assigned as a light-flavour jet. The matching is exclusive, which means that each particle is only assigned to one jet. **Add a reference, possibly performance** This definition of truth flavour in simulation is used generally within this thesis.

### 3.2.2 Algorithm descriptions

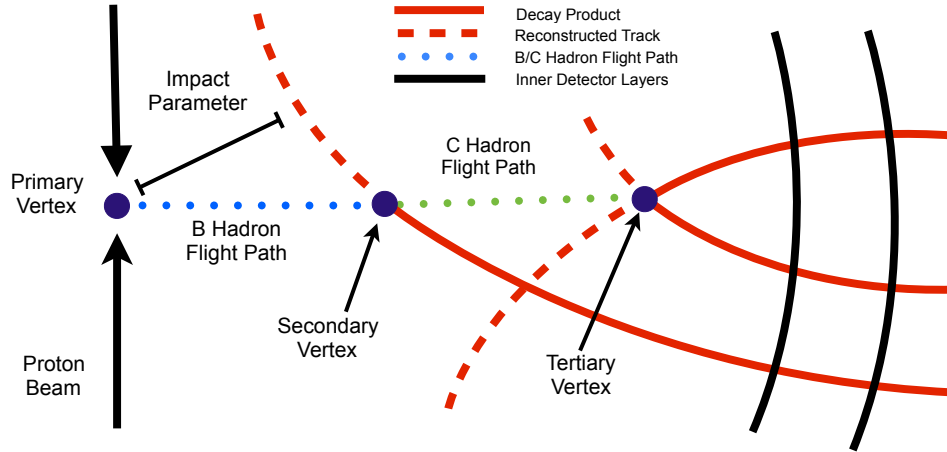
To identify  $b$ -jets,  $b$ -tagging algorithms utilise the long lifetimes of the heavy-hadrons that decay through the flavour changing weak interaction, a hadron containing a  $b$ -quark has a lifetime of the order 1.5 ps. A  $b$ -jet decay chain will typically contain two of these flavour

---

<sup>1</sup>Section ?? contains a concrete example of this.

changing interactions, as at the quark level, the  $b$ -quark contained in the jet will decay to a  $c$ -quark, which will then decay into a  $u$  or  $d$  quark. The long lifetimes of the heavy flavour hadrons means that they will decay a measurable distance from the primary vertex, the point where the hard-scatter collision occurs. Hence, the flavour of a jet can be inferred from the presence of particles that originate from a point offset from the primary vertex. In practice this is performed using the topology of tracks formed by the inner detector and measurements from the calorimeter, which have been described in Section 2.2.2 and 2.2.3 respectively.

There are three base  $b$ -tagging algorithms utilised to produce discriminating variables [?], which are described in the next three sections. The variables are then combined in a multi-variate algorithm which is described in Section 3.2.2.4. Figure 3.3 shows a schematic illustrating how the tracks from the inner detector are used by the three  $b$ -tagging algorithms to identify a  $b$ -jet, the details of this figure are referred to in the following three sections.



**Figure 3.3:** A diagram to illustrate the key features of a  $b$ -jet that are utilised by the base  $b$ -tagging algorithms.

### 3.2.2.1 Impact parameter based

The IP3D algorithm utilises the impact parameter, which is defined as the shortest distance between a specific track and the primary vertex. A track corresponding to a particle coming from the offset decay vertex of a heavy-hadron is likely to have a large impact parameter, meaning that the distribution of track impact parameter is different for each of the jet-flavours. The impact parameter of a track coming from the decay of a heavy hadron is

indicated in Figure 3.3. In this algorithm, for all tracks associated to a jet, the impact parameter is calculated in both the transverse (perpendicular to beam-line) and longitudinal (parallel to beam-line) direction, which are referred to as  $d_0$  and  $z_0$ . Then the IP3D algorithm calculates a likelihood of the jet having a specific flavour, based on the distributions of the impact parameters ( $d_0$ ,  $z_0$ ) and their significances ( $d_0/\sigma_{d0}$  and  $z_0/\sigma_{z0}$ ) for tracks within the jet. Another similar algorithm, IP2D, also calculates the jet flavour likelihood from just the transverse distributions, ( $d_0$  and  $d_0$  significance), which is more robust to pile-up, as tracks from pile-up jets are likely to have a large  $z_0$  significance value.

### 3.2.2.2 Secondary vertex

The SV1 algorithm aims to reconstruct a secondary vertex of two or more intersecting tracks, corresponding to the decay of a heavy-flavour hadron; the secondary vertex within a  $b$ -jet's decay chain is illustrated in Figure 3.3. The SV1 algorithm then calculates many variables that are associated with properties of the reconstructed secondary vertex that show flavour discrimination; some example variables are the vertex invariant mass, which will be larger for  $b$ -jets due to the heavy mass of the  $b$ -hadron<sup>2</sup>, the distance in the transverse plane between the primary vertex and the secondary vertex, which will be larger for  $b$ -jets due to the long lifetime of the  $b$ -hadron, and the number of tracks at the secondary vertex, which will be larger for reliable secondary vertices.

### 3.2.2.3 Jet Fitter

The JetFitter algorithm (JF) attempts to reconstruct the full decay chain of the  $b$ -hadron into a charmed-hadron and then into light-hadrons. This is done by assuming that all vertices lie on a common  $b$ -flight axis, and then constructing vertices from the intersection of one or more tracks and the flight axis. The aim of this is to reconstruct the secondary and tertiary vertices which correspond to the decays of the  $b$ -hadron and charmed-hadron, as illustrated in Figure 3.3. Similar to SV1, the JetFitter algorithm then calculates a number of flavour discriminating variables: for example vertex mass and number of vertices with two or more tracks.

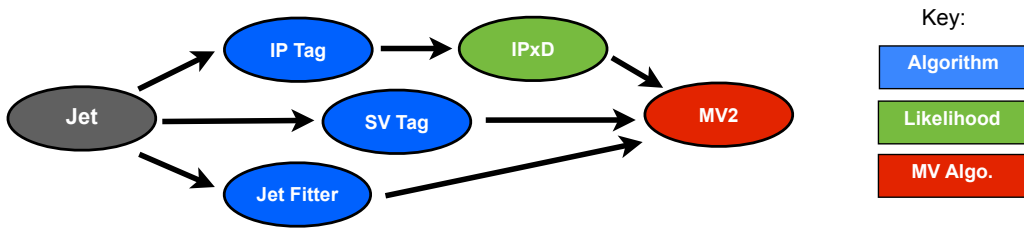
---

<sup>2</sup>Mass of a B-meson  $\sim 5$  GeV and the mass of a D-meson  $\sim 1.9$  GeV, which are the most common heavy hadrons in a  $b$ - and  $c$ -jet respectively.



### 3.2.2.4 Multi-variate

The three base algorithms are combined in a boosted decision tree (BDT), a machine-learning technique for combining the many flavour-discriminating variables, resulting in an algorithm that is known in MV2. As shown in Figure 3.1, MV2 combines the likelihood output of IP3D and IP2D with the discriminating variables of SV1 and JF discussed in the preceding sections, resulting in an output between -1 and 1, where 1 indicates that the jet is likely to be a  $b$ -jet and -1 indicates the inverse.



**Figure 3.4:** A diagram illustrating how three base flavour tagging algorithms are combined in the MV2 algorithm.

A cut is then applied to this MV2 output in order to select jets that are likely to  $b$ -jets. The cut can be varied to create different operating points which vary the  $b$ -jet efficiency, light-jet rejection and  $c$ -jet rejection<sup>3</sup>. Looser operating points have a relatively low cut on MV2, meaning that the  $b$ -jet efficiency is higher at the cost of worse light- and  $c$ -jet rejections, and the inverse is true for tighter operating points. Table ?? shows the list of fixed cut operating points that are used in ATLAS with a given cut on MV2 output; shown with the corresponding benchmark  $b$ -jet efficiency,  $c$ -jet rejection, light-jet rejection and  $\tau$  rejection. For the remainder of this thesis, the operating points will be referred<sup>4</sup>

| MV2 Cut Value | $b$ -jet efficiency [%] | $c$ -jet rejection | Light-jet rejection | $\tau$ rejection |
|---------------|-------------------------|--------------------|---------------------|------------------|
| 0.9349        | 60                      | 34                 | 1538                | 184              |
| 0.8244        | 70                      | 12                 | 381                 | 55               |
| 0.6459        | 77                      | 6                  | 134                 | 22               |
| 0.1758        | 85                      | 3.1                | 33                  | 8.2              |

**Table 3.1:** The Mv2c10  $b$ -tagging algorithm operating points; with the corresponding  $b$ -jet efficiency,  $c$ -jet rejection, light-jet rejection and  $\tau$  rejection. This table is taken from reference [?].

<sup>3</sup> $b$ -jet efficiency is defined as the probability of accepting a true  $b$ -jet. Light-jet rejection is defined as 1 divided by the probability of accepting a true light-jet. A similar definition applies for  $c$ -jet rejection.

<sup>4</sup>In this thesis only the fixed-cut operating points shown above will be used, however, there also exists a set of flat efficiency operating points where the MV2 cut depends on jet- $p_T$

The BDT is trained using a simulated sample of  $t\bar{t}$  events that will contain a mix of  $b$ -,  $c$ - and light-jets as well as a sample containing  $Z'$  boson decaying to  $b$ -quarks. The training makes use of the truth flavour labels assigned to jets using the process described in Section 3.2.1. A training sample with known truth labels is required as this allows the BDT to be optimised such that it uses the complex correlations between the input variables to allow for high  $b$ -jet efficiencies whilst still obtaining a large  $c$ - and light-jet rejection. The recommended  $b$ -tagging tool is MV2c10 which has been trained on a sample containing 10% charm-jets, to give strong light- and  $c$ -rejection.

### 3.2.3 Performance

I have a couple of options here. =<sub>i</sub> Put performance in event selection (tagging efficiency plot) =<sub>i</sub> Describe performance here with some discussion of high pT b-Tagging

## 3.3 **b**Tagging: Validation in Dijet Events

## 3.4 Leptons

### 3.4.1 Electron

### 3.4.2 Muon

## Chapter 4

# Triggering in the di- $b$ -jet analysis

As described in Section 2.3, ATLAS does not have the resources to process and store all the data from the 40 MHz of collisions delivered by the LHC. To solve this problem the ATLAS trigger system performs the vital role of reducing the rate of data-taking to 1 kHz by selecting events containing a high- $p_T$  object.

As a result all analyses must choose a trigger strategy and understand the impact of this trigger on their analysis. In the di- $b$ -jet analysis a single jet trigger is used for the high-mass channel and a double  $b$ -jet trigger for the low-mass channel. This chapter aims to provide a detailed description of the triggers used in this analysis, and as such is organised in the following manner; Section 4.1 provides a brief description of jet triggers as used in the high-mass channel and the limitations of this approach, Section 4.2 contains a description of  $b$ -jet triggers that are used in the low-mass channel and finally Section 4.3 presents the measurement of the  $b$ -jet trigger efficiency, an essential input of the low-mass channel.

## 4.1 Jet-Triggers

Jet-triggers are tasked with selecting events with one or more jets from the deposits in the ATLAS calorimeter system, this is one of most challenging triggers in any hadron-hadron collider due to the extremely high cross-sections of hadronic jet production [23]. In Run-2 the jet-triggers are used at both L1 and HLT level; each using different levels of information and different algorithms, so are described separately within this section.

### 4.1.1 Level 1

The L1 trigger is a hardware based trigger which accepts or rejects an event within  $2.2\ \mu\text{s}$ . The L1 jet-trigger receives trigger towers from the calorimeter; where a trigger tower is the measured energy deposit in a cell of the ECAL or HCAL of granularity  $0.1 \times 0.1$  in the  $\eta - \phi$  plane. In the L1 trigger hadronic jet algorithms search for a neighbouring group of  $4 \times 4$  trigger towers containing energy deposits above some pre-set threshold. Our analysis uses the L1 trigger known as L1\_J100, which requires that at least one trigger tower group with an energy deposit of 100 GeV has been found. Other L1 triggers that search for multiple clusters are also possible to reduce the energy thresholds required. The L1 trigger then seeds the HLT trigger. It is also worth noting that at L1 there is no tracking information, meaning that electron and taus are also triggered on using similar techniques as hadronic jet algorithms, except using narrower groups of trigger towers.

### 4.1.2 HLT

The HLT trigger is a software based trigger which, due to the lower input rate and larger time window, is able to use more complex algorithms to reconstruct jets. At the HLT level jets are reconstructed using topoclusters (TCs) constructed from neighbouring cells selected using the cell's energy significance ( $E/\sigma$ ); TCs are seeded from cells with  $E/\sigma > 4$ , then neighbouring cells with  $E/\sigma > 2$  are added and finally all neighbouring cells around are also added. Jets are then reconstructed from the topoclusters; in this analysis jets have been reconstructed using the anti- $k_T$  algorithm with an  $R = 0.4$ <sup>1</sup>.

### 4.1.3 High-mass trigger selection

For the high-mass analysis the trigger HLT\_j380 is used, that is fired when a jet is found with a  $p_T > 380$  GeV. This is chosen as it is the lowest un-prescaled single jet-trigger; meaning that of triggers that accept every event passing a single jet criteria, this trigger has the lowest jet- $p_T$  threshold. Due to the exponential increase in jet production cross-section at low jet- $p_T$ , the  $p_T$  threshold is set to keep the acceptance rates low enough such that the HLT trigger is within its output rate budget of 1 kHz.

However, as will be discussed further in Section ??(sec:evtSel), this  $p_T$  threshold limits

---

<sup>1</sup>Section 3.1 (sec:obj-jets) defines these terms

the high-mass di-*b*-jet analysis to only select events with  $m_{jj} > 1.2$  TeV. Otherwise the  $m_{jj}$  range will enter a kinematic region where trigger acceptance is less than 1 in such a way that the QCD background is sculpted in a manner that the background modelling can not adapt to. To reach lower masses a different trigger strategy is required.

## 4.2 *b*-Jet Triggers

This analysis searches for pairs of *b*-jets, which, as described in Section 3.2(*sec:obj-bjets*), can be identified from the topology of tracks in the inner detector indicating that a *B*-hadron was within the jet. The same techniques can be used at the trigger level to reduce rates significantly <sup>2</sup> allowing a lower jet- $p_T$  threshold than was used by the single jet- $p_T$  trigger, and hence lower  $m_{jj}$  values to be accessed. *b*-jet triggers have been used in a range of previous ATLAS analyses including for searches for exotic particle decaying into a pair of Higgs bosons, which then decay to 4 *b*-jets [24].

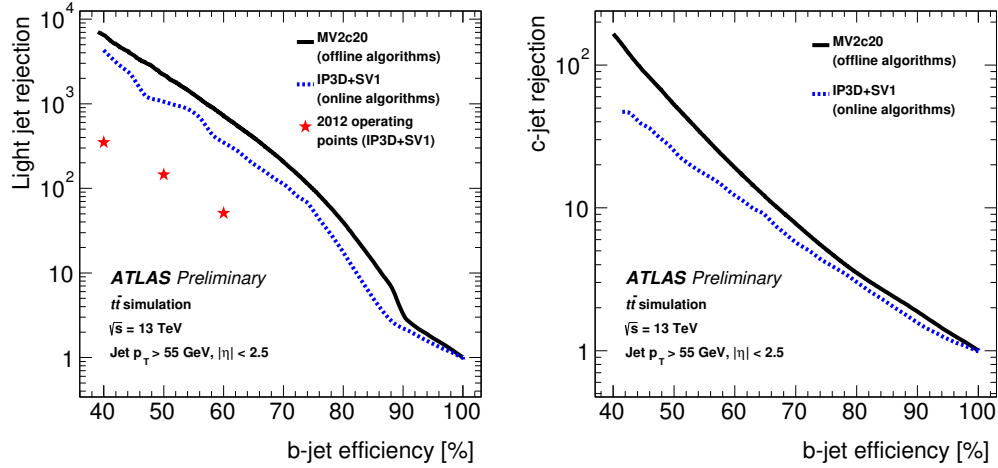
### 4.2.1 General description

In 2016 data, the *b*-jet trigger configuration contains three steps [25], making use of the regions of interest (RoI) described by the jets found by the jet-trigger. Firstly, a ‘fast’-tracking algorithm is run in a super-RoI which is formed around all jets in the event which have  $E_T > 30$  GeV; these tracks are then used to identify the primary vertex in the event. Secondly, within each jet RoI precision tracking is run, with a constraint on the PV position from the first step. Finally, these tracks are the input to the multi-variate *b*-tagging algorithm described in Section ??(*sec:obj-bjets\_MV2*) to identify *b*-jets. There are several *b*-jet triggers available in the ATLAS trigger menu; with a variety of requirements on the jet multiplicity, number of tagged jets and *b*-tag operating point used. Figure 4.1 shows ROC curves representing the expected performance of the Run-2 *b*-jet trigger.

There are few subtleties worth commenting on the *b*-jet trigger configuration which affect decisions taken in this analysis. One is that on this figure there are two lines corresponding to different *b*-tagging algorithms used in *b*-jet trigger; IP3D+SV1 was used in 2015 data-taking, whilst the MV2c20 was used in 2016 data-taking. Another difference between 2015 and 2016 is the primary vertex finding algorithm used; 2016 data-taking em-

---

<sup>2</sup>It is known that the QCD background is dominated by light-jets, see Figure ??*Plot of background flav comp*



**Figure 4.1:** The expected *b*-jet efficiency of *b*-jet triggers with respect to (a) light-jet and (b) *c*-jet rejection in the case where the *b*-tagging algorithm used is MV2c20 (Black), IP3D+SV1 (Blue) and for the set-up used in Run-1 (red stars) [25].

ployed an algorithm based on offline primary vertex finding, known as `xPrmVtx`, whilst in 2015 an algorithm using a simpler histogram based approach was employed, known as `EFHist`.

Finally it is worth noting that there are differences between online and offline *b*-tagging that will have an impact on what is to follow. Firstly, coarser tracking information is available online, notably online tracks are not reconstructed from the whole range of the detector. Secondly, a slightly different training setup is used for the multi-variate algorithm, mainly that a different fraction of *c*-jets were present in the training sample (10% offline vs. 20% online).

In this analysis a double *b*-jet trigger is used,

`HLT_j150_bmv2c2060_split_j50_bmv2c2060_split`

which triggers on two jets with  $p_T > 150$  and 50 GeV respectively, which have been *b*-tagged at the 60% efficiency working point.

### 4.3 Efficiency Measurement of the $b$ -Jet Trigger

Any part of the ATLAS detector framework needs to be understood and calibrated with data for use in an analysis; and this includes the trigger which can have a large impact on the analysis. In this section I discuss the strategy and results of the  $b$ -jet trigger efficiency measurement in 2016, which is an important input to the low-mass channel of the di- $b$ -jet analysis.

For clarity in this section I would like to make two definitions clear. Online refers to any algorithms run or objects reconstructed at the trigger level, offline refers to algorithms run after events have passed the trigger at the data-processing level.

#### 4.3.1 Strategy

The  $b$ -jet trigger is always used in tandem with offline  $b$ -tagging which is calibrated independently of the  $b$ -trigger. As mentioned before, there are many differences between offline and online  $b$ -tagging. Hence, to do this measurement whilst making use of the offline  $b$ -tagging calibrations already available,  $b$ -jet trigger efficiency with respect to offline  $b$ -tagging,  $\epsilon_{bTrig}$ , is measured. This is defined as the number of offline-tagged true  $b$ -jets that match an online-tagged trigger-jet divided by the number of offline tagged  $b$ -jets that match a trigger jet. Or to put this in an equation;

$$\epsilon_{bTrig} = \frac{N(\text{Offline-tagged, online-tagged, true } b\text{-jets})}{N(\text{Offline-tagged, trigger-matched, true } b\text{-jets})} \quad (4.1)$$

This quantity can be interpreted as the probability that a true  $b$ -jet is tagged at the trigger-level, given that there is a jet at the trigger level and that it would be  $b$ -tagged at the offline stage.

To measure  $\epsilon_{bTrig}$  a sample that has high  $b$ -jet purity is required, such that jets used to calculate this ratio are true  $b$ -jets. It is also necessary to trigger on this sample in such a way that there is no bias from using  $b$ -tagging online; or simply put the  $b$ -jet trigger cannot be used to select events. The sample used to fill these criteria is a di-lepton  $t\bar{t}$  sample containing a muon and an electron. Top-quarks decay to a  $W$ -boson and a  $b$ -quark with almost 100% branching ratio meaning that this sample provides a good source of  $b$ -quarks, but also the electron and muon give a distinct signature which allows us to select this process with good

purity and gives a non- $b$ -jet object to trigger on. The exact event selection is described below.

The  $b$ -jet trigger efficiency is determined in data and is compared to the efficiency found in a simulated  $t\bar{t}$  sample which is used to extrapolate the efficiency to higher jet- $p_T$  where the data-derived efficiency loses statistical precision. The efficiency in data, including the simulation based extrapolation, can then be compared to simulation to derive a Data/Monte-Carlo scale factor, which is used as the input to the analysis.

$\epsilon_{bTrig}$  and Data/Monte-Carlo scale factors are derived for all combinations of offline and online  $b$ -tagging working points. However, only the process for the 70% offline and 60% online working point is shown as this is set of working points used in this analysis.

### 4.3.2 Datasets

The data used for this analysis is the full 2016 ATLAS data-set. In addition to the usual data-quality requirements applied, as discussed in Section ??(sec:evtSel\_GRL), a  $b$ -jet trigger aware Good Run List (GRL) <sup>3</sup> applies the requirement that the online beamspot  $z$ -position is within 2mm of the origin in Periods A-I of the data. This means that the data-set contains  $24.5 \text{ fb}^{-1}$  of data. A discussion of the requirement for this GRL is in Section 4.3.6.

For the simulated  $t\bar{t}$  sample, the generation is performed with a Powheg-Box v2 [26] generator with the CT10 PDF sets in the matrix element calculations. Also considered is a simulated single-top sample; electroweak t-channel, s-channel and  $Wt$ -channel single top-quark events are generated using the Powheg-Box v1 generator. This generator uses the 4-flavour scheme for the NLO matrix elements calculations together with the fixed four-flavour PDF set CT10f4. For both processes the parton shower, fragmentation, and the underlying event are simulated using Pythia6.428 [27] with the CTEQ6L1 [28] PDF sets and the corresponding Perugia 2012 tune (P2012) [29]. The top mass is set to 172.5 GeV. The EvtGen v1.2.0 program [30] is used for properties of the bottom and charm hadron decays.

---

<sup>3</sup>A GRL is effectively a list of lumi-blocks that pass certain data-quality requirements. As mentioned in the text a further discussion is held here in Section ??(sec:evtSel\_GRL)



### 4.3.3 Event Selection

A high-purity sample of  $b$ -jets is selected using a di-lepton  $t\bar{t}$  selection.

The event selection is summarised as follows:

- The event fired a single lepton bperf trigger which are:
  - HLT\_mu26\_imedium\_2j35\_bperf
  - HLT\_e26\_tight\_iloose\_2j35\_bperf
  - HLT\_e26\_lhtight\_iloose\_2j35\_bperf
- At least 1 medium muon:  $p_T > 25$  GeV, which has no jet within a  $\Delta R$  of 0.4.
- At least 1 medium electron:  $p_T > 25$  GeV.
- 2 offline  $b$ -tagged jets, defined as:
  - Offline  $R=0.4$  anti- $k_T$  jets.
  - $p_T > 35$  GeV and  $|\eta| < 2.5$ .
  - Offline  $b$ -tagged at the 85% operating point.
  - Jet must be matched to a trigger-jet.

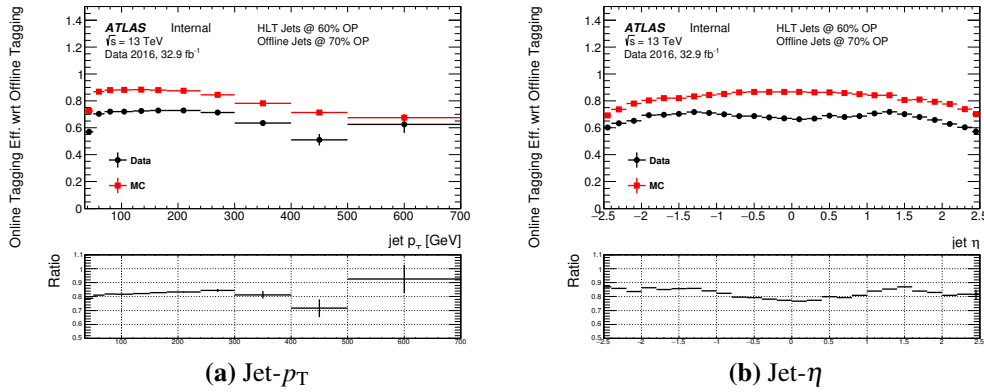
Descriptions of the object-definitions of muons, electrons, jets and  $b$ -tagged can be found in Sections 3.4.2(*sec:obj-muon*), 3.4.1(*sec:obj-elec*), 3.1(*sec:obj-jet*) and 3.2(*sec:obj-bjet*) respectively. Online trigger jets are matched exclusively to offline jets using  $\Delta R$  matching, requiring for a match the jets must have  $\Delta R < 0.6$ .

The triggers used are bperf trigger, which are special triggers used in data-taking specifically for monitoring the  $b$ -jet trigger performance. They fire if a muon or an electron with  $p_T > 26$  GeV is reconstructed at the trigger level. The bperf triggers then run the online  $b$ -tagging algorithm on all trigger jets with  $|\eta| < 2.5$  and  $p_T > 35$  GeV without performing any cuts on the output of the multi-variate algorithm; ensuring there is no bias in the efficiency measurement.

#### 4.3.4 The Initial Problem

To give context to the following section; the first discussion will be what was first observed when measuring the  $b$ -jet efficiency. To show the problems observed clearly, in this section the initial event selection is replicated; hence no  $b$ -jet trigger aware GRL is applied, offline jets are not required to match a trigger jet in the denominator and the triggers required are single lepton triggers without the additional  $b$ -perf functionality<sup>4</sup>. In addition, for this and the following two sections simulation refers to  $t\bar{t}$  only, but it will be shown later that the effect of single-top production is small so the conclusions here are still valid.

Figure 4.2 shows  $\varepsilon_{bTrig}$  against jet- $p_T$  and jet- $\eta$ ; the efficiency in data is substantially below the efficiency expected from simulation and shows a clear shape in jet- $\eta$  distributions. This substantial differences need to be investigated and understood.



**Figure 4.2:** The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for Data (black) and simulation (red) against jet- $p_T$  (a) and jet- $\eta$  (b). The  $b$ -trigger aware GRL is not applied and trigger matching is not required.

#### 4.3.5 Investigation

Given the disagreements between data and simulation shown above a number of cross-checks were performed to understand this discrepancy, including checking for performance dependence on period, detector performance, pile-up conditions and online beamspot position. In this section, I summarise the results of the investigation and our understanding of the  $b$ -trigger performance in 2016 data. For this, the set-up as described in 4.3.3 is used

<sup>4</sup>Specifically HLT\_mu26\_imedium, HLT\_e26\_tight\_loose and HLT\_e26\_lhtight\_loose.

with the exception that the  $b$ -jet aware GRL is not applied to allow us to see the problems clearly.

The major problem that was discovered to be causing the large discrepancies was related to primary vertex finding. As described above, in 2016 data an algorithm known as `xPrmVtx` was used to find the primary vertex. It has since been uncovered that there was a bug in the code used to implement this algorithm; effectively different co-ordinates were used by different components of the code. Online tracks passed to `xPrmVtx` use position with respect to online beam-spot position, where the `xPrmVtx` algorithm assumed track position with respect to the origin. This means that when the online beamspot  $z$ -position is far from the origin, a dummy vertex with position at the origin is passed to the  $b$ -tagging algorithms. This leads to sub-optimal performance, as will be shown below. For ease of reading online beamspot  $z$ -position is henceforth referred to as  $z_{bs}^{online}$ .

The exact setup for the  $b$ -jet trigger has changed as data has been taken, to respond to performance issues as they are noticed and patches are applied. As such the relevant conditions of the  $b$ -jet trigger can be split into three regions of data-taking, which I will refer to as epochs. The effect of `xPrmVtx` returning a dummy vertex on  $b$ -jet trigger performance is different in each of these epochs, the details are summarised Table 4.1. As a result of these differences in trigger performance, each epoch is now considered independently.

| Epoch | Runs                  | Periods               | Effect if no <code>xPrmVtx</code> PV is found          |
|-------|-----------------------|-----------------------|--|
| 1     | 296939-300571, 300655 | A,B(part)             | An invalid vertex is passed to the online $b$ -tagging |
| 2     | 300600, 300784-308084 | B(part),C,D,E,F,G,I,J | The $b$ -jet trigger is not fired                      |
| 3     | 309331-311481         | K,L                   | A back-up primary vertex finding algorithm is used.    |

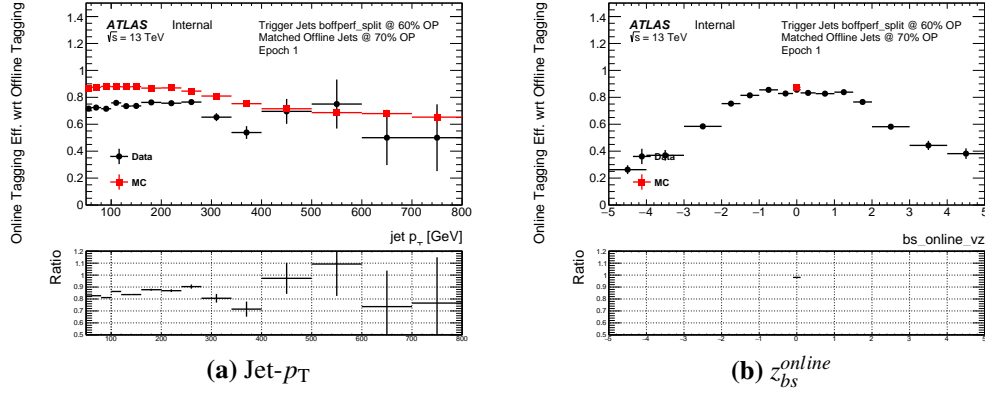
**Table 4.1:** A table describing the effect of not finding a valid `xPrmVtx` primary vertex on different epochs of data.

Firstly let us consider Epoch 1; Figure 4.3(a) shows that efficiency against  $\text{jet-}p_T$  is 80-90% of that in simulation, similar to that shown in the previous section. However, Fig-

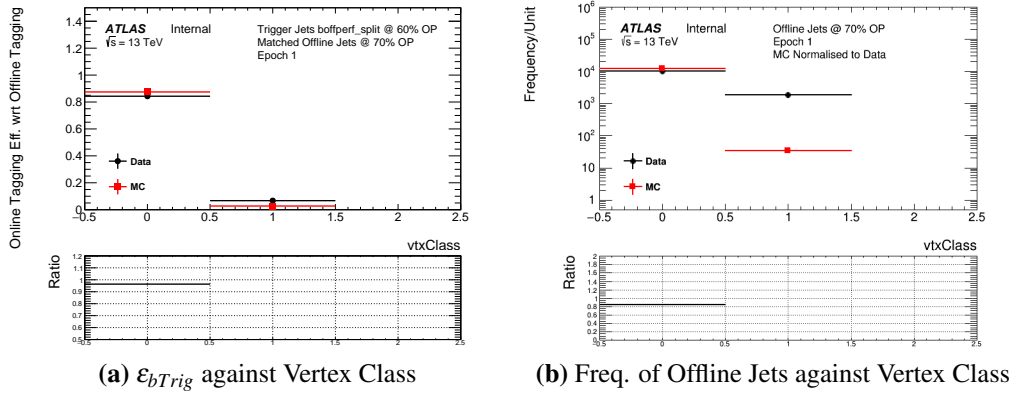
Figure 4.3(b) shows that  $\epsilon_{bTrig}$  in Epoch 1 has a strong dependence of  $z_{bs}^{online}$ ; when  $z_{bs}^{online}$  is close to zero  $\epsilon_{bTrig}$  in data and simulation are comparable <sup>5</sup> but as  $|z_{bs}^{online}|$  increases efficiency falls off steeply. To understand this performance the variable ‘vertex class’ is studied, which is defined as 0 when a valid `xPrmVtx` vertex is found and 1 if not. Figure 4.4(a) shows that when an `xPrmVtx` vertex is found  $\epsilon_{bTrig}$  is reasonably high ( $\sim 0.8$ ) and is comparable between data and simulation (within 5%), whilst if no valid `xPrmVtx` vertex is found then efficiency is very low in both simulation and data. However, Figure 4.4(b) shows that a valid `xPrmVtx` vertex is found in simulation in  $> 99\%$  of the jets, whilst in data there is  $\sim 16\%$  of events where no valid `xPrmVtx` vertex is found. Hence, combining the information in Table 4.1, Figure 4.3 and Figure 4.4 it can be concluded that in Epoch 1 in events where the  $|z_{bs}^{online}|$  is far from 0 then `xPrmVtx` returns a dummy vertex which results in a low  $\epsilon_{bTrig}$ , explaining the data/simulation differences in Epoch 1.

---

<sup>5</sup>In simulation the  $z_{bs}^{online}$  is always set to zero.



**Figure 4.3:** The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for data from Epoch 1 (black) and simulation (red) against jet- $p_T$  (a) and online beamspot  $z$ -position (b). The  $b$ -jet trigger aware GRL has not been applied.



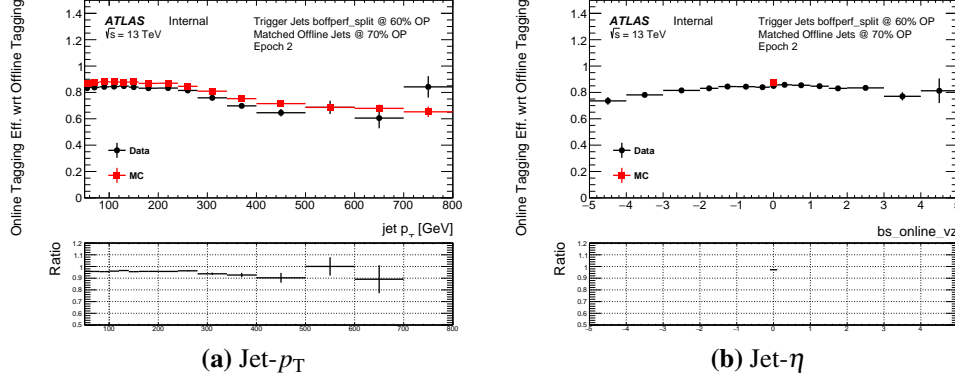
**Figure 4.4:** (a) The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag and (b) the number of offline jets passing 70% operating point tag and matching a HLT trigger jet against vertex class for data from Epoch 1 (black) and simulation (red). The  $b$ -jet trigger aware GRL has not been applied.

In Epoch 2, there is a similar problem to Epoch 1, but there is a subtle difference which requires us to look at this region in a different way. As in Epoch 1, when  $z_{bs}^{online}$  is far from zero then a  $\times P r m V t \times$  PV is not found. However in Epoch 2 this means that the  $b$ -jet trigger was discovered to falsely terminate whilst processing the event, meaning that there are no online  $b$ -jets are available in the event, and therefore the trigger will not fire. However, the additional complication compared to Epoch 1 is this means that the  $b$ -perf triggers used to measure the efficiency are also not fired when no valid  $\times P r m V t \times$  PV is available. Hence, measuring  $\epsilon_{bTrig}$  using the set-up as described will not capture the cases where a valid  $\times P r m V t \times$  PV is found and thus  $\epsilon_{bTrig}$  should be consistent in data and simulation; Figure 4.5 shows that the  $\epsilon_{bTrig}$  measured in data to be in agreement with

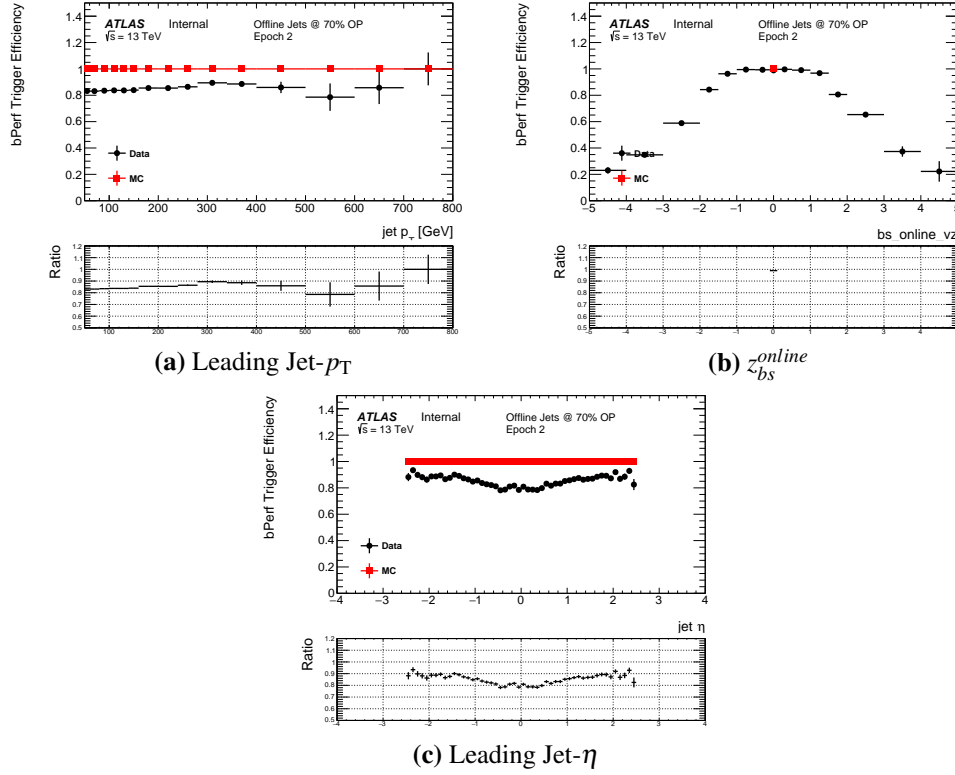
simulation within 5%.

For Epoch 2, in addition to measuring  $\epsilon_{bTrig}$  it is necessary to also account for the cases when a false  $xPrmVtx$  PV is found. This is done by measuring the  $b$ -perf efficiency,  $\epsilon_{bPerf}$ , the efficiency that there is a valid primary vertex in the event.  $\epsilon_{bPerf}$  is calculated by dividing the number of events that pass the trigger `HLT_mu26_imedium_2j35_bperf` by the number that pass the trigger `HLT_mu26_imedium`, such that the denominator has no  $b$ -trigger dependency so is unaffected by  $xPrmVtx$  PV. This is an event level quantity and as such is measured with respect to other event level quantities, such as leading jet- $p_T$ . Figure 4.6 shows that:  $\epsilon_{bPerf}$  has a data/simulation ratio of around 80% which is similar to that in Section 4.3.4 and  $\epsilon_{bPerf}$  shows similar behaviour with respect to  $z_{bs}^{online}$  as observed in Epoch 1. Finally it is observed that  $\epsilon_{bPerf}$  has a lower efficiency at smaller values of absolute leading jet- $\eta$ ; this is due to the fact that at high- $\eta$  tracks have a larger error on the longitudinal impact parameter,  $z_0$ , meaning that the mis-match of co-ordinates can in the  $xPrmVtx$  algorithm is covered by the errors, mitigating this issue. This effect must be accounted for in the final efficiency measurement. **This last two sentences are dodgy**

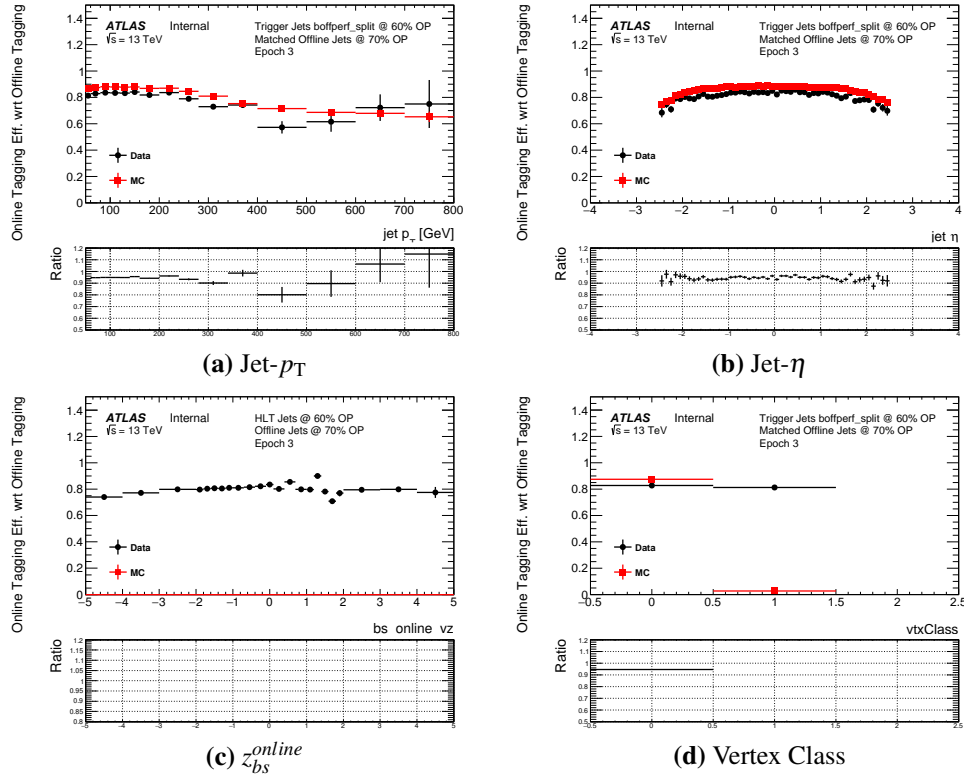
Epoch 3, when no  $xPrmVtx$  PV is found then a backup PV finding algorithm is used, known as `EFHist`, which finds the PV through a basic histogramming of the tracks, the simplicity of the algorithm means that a PV can be found as long as 1 track is present. Figure 4.7 shows  $\epsilon_{bTrig}$  for Epoch 3 for jet- $p_T$ , jet- $\eta$ ,  $z_{bs}^{online}$  and vertex class (as defined above). In Epoch 3  $\epsilon_{bTrig}$  measured in data is within 5% of simulation and there is no shape difference between the two with respect to jet- $\eta$ . In addition it is shown that in Epoch 3 there is no strong dependence on  $z_{bs}^{online}$ , and that efficiency in data is consistent if a valid  $xPrmVtx$  vertex or not (vertex class = 0 or 1 respectively). This demonstrates the success of the backup vertex approach.



**Figure 4.5:** The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for data from epoch 2 (black) and simulation (red) against jet- $p_T$  (a), jet- $\eta$  (b) and online beamspot  $z$ -position (c).



**Figure 4.6:**  $b$ -perf efficiency,  $\epsilon_{bPerf}$ , for data from Epoch 2 (black) and simulation (red) against leading-jet  $p_T$  (a), online beamspot  $z$ -position (b) and leading jet- $\eta$  (c).



**Figure 4.7:** The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for data from Epoch 3 (black) and simulation (red) against (a) jet- $p_T$ , (b) jet- $\eta$ , (c) online beamspot  $z$ -position and (d) vertex class.



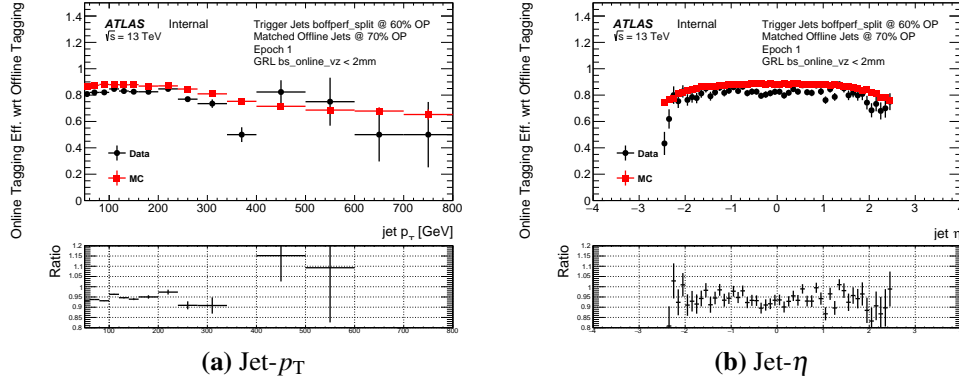
#### 4.3.6 Solution: $b$ -Jet Trigger GRL

To summarise, in the previous section it is shown that at large values of absolute online beamspot  $z$ -position the measured  $\epsilon_{bTrig}$  in Epoch 1 and  $\epsilon_{bPerf}$  in Epoch 2 is lower in data than in MC, due to poor `xPrmVtx` PV finding performance. In Epoch 3 there is reasonable data/simulation agreement due to the use of a backup vertex finding algorithm.

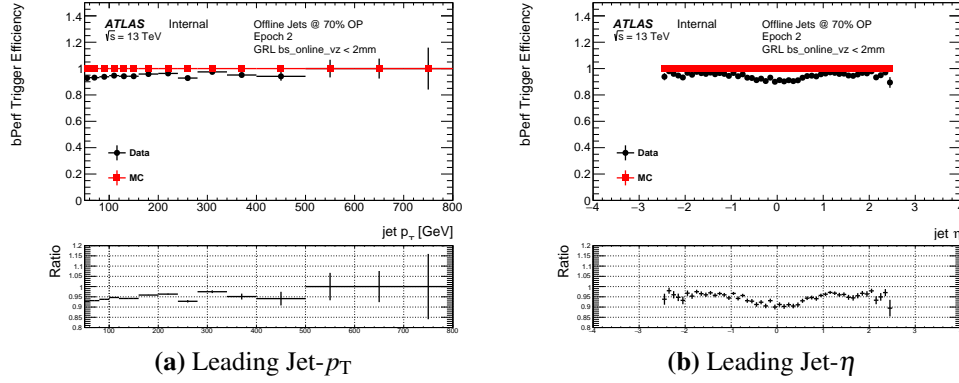
The solution employed is to apply a  $b$ -jet trigger aware GRL that removes events with an absolute  $z_{bs}^{online}$  greater than 2mm in Epoch 1 and 2, such that the events with low efficiency are removed. The cost of this approach is the luminosity of our data-set is reduced, specifically the data-set falls from  $32.9 \text{ fb}^{-1}$  to  $24.3 \text{ fb}^{-1}$ . However there are three key reasons why use of a  $b$ -jet trigger GRL was chosen over simply applying an overall efficiency. Firstly, as there is no beamspot position distribution in simulation it is not clear that kinematics of events at high  $z_{bs}^{online}$  can be well understood and modelled; the sculpting of the efficiency with respect to jet- $\eta$  is an example of this. Secondly, the efficiencies are quite low at high beamspot  $z$ -position, so the loss in luminosity  $\times$  acceptance is relatively small and finally the use of a GRL means a more realistic estimate of the actual luminosity used in an analysis is used.

For the choice of which value of beamspot  $z$  position to use for in the GRL, it was required to select the widest cut where the efficiency had not significantly declined, such that as much luminosity as possible is retained while removing most of the affected region. This 2 mm cut was chosen from examining Figure 4.3(b) and Figure 4.6(b) and from studying a variety of cuts from 2 mm to 1 mm.

After the GRL is applied,  $\epsilon_{bTrig}$  for Region 1 becomes approximately 90-95% of the efficiency measured in simulation, as shown in Figure 4.8, and  $\epsilon_{bPerf}$  for Region 2 becomes approximately 95% of the efficiency measured in simulation, as shown in Figure 4.9.

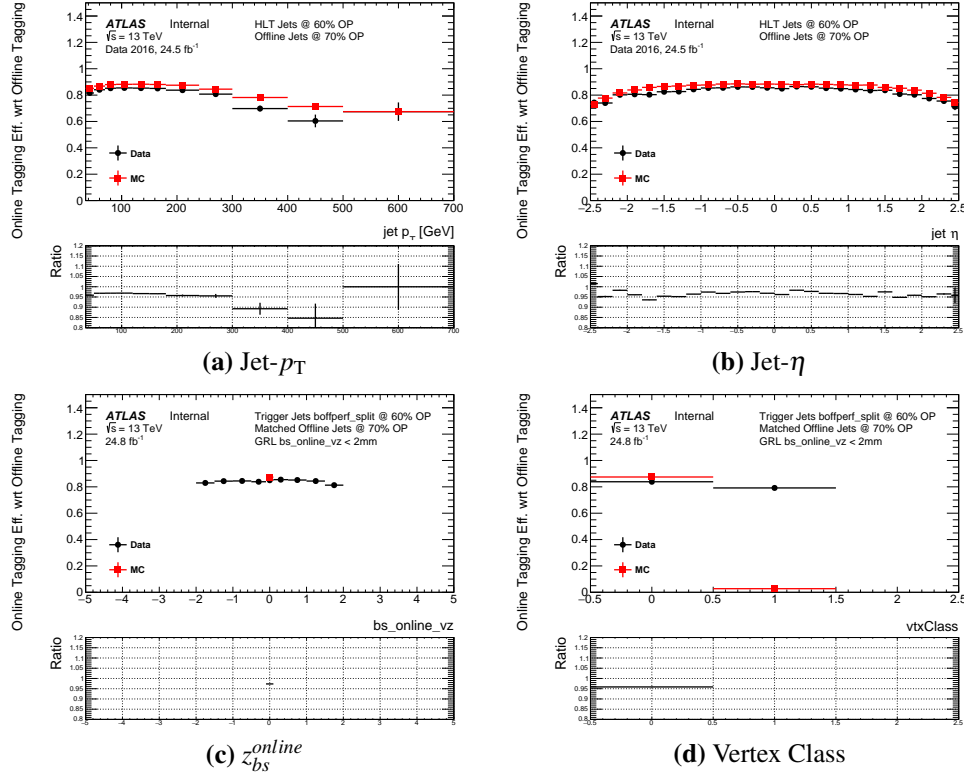


**Figure 4.8:** The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for data from Region 1 (black) and simulation (red) against  $\text{jet-}p_T$  (a) and  $\text{jet-}\eta$  (b). The  $b$ -jet trigger aware GRL has been applied.

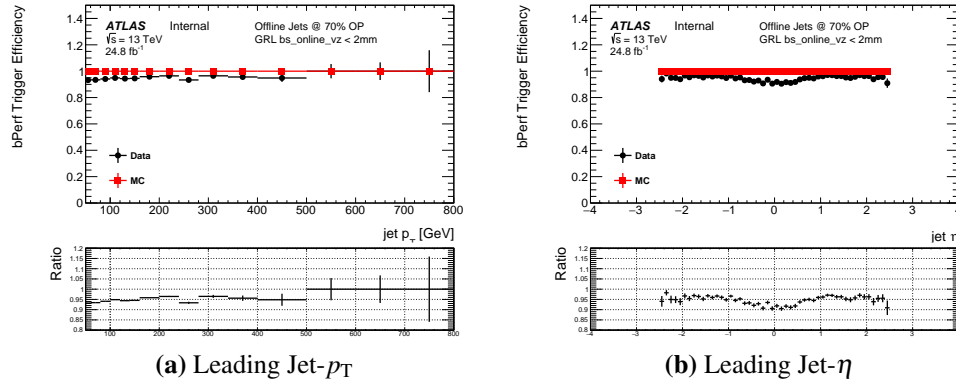


**Figure 4.9:**  $b$ -perf efficiency,  $\epsilon_{bPerf}$ , for data from Region 2 (black) and simulation (red) against leading (a)  $\text{jet-}p_T$  and (b)  $\text{jet-}\eta$ . The  $b$ -jet trigger aware GRL has been applied.

Figures 4.10 and 4.11 shows measured  $\epsilon_{bPerf}$  and  $\epsilon_{bTrig}$  for the full 2016 data-set, combining Regions 1, 2 and 3, with the  $b$ -jet trigger aware GRL applied. This represents the raw observed data/simulation efficiencies when the full event selection has been applied.



**Figure 4.10:** The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for the full 2016 data-set (black) and simulation (red) against jet- $p_T$  (a), jet- $\eta$  (b), online beamspot  $z$ -position (c) and vertex class (d).



**Figure 4.11:**  $b$ -perf efficiency,  $\epsilon_{bPerf}$ , for the full 2016 data-set (black) and simulation (red) against (a) leading jet- $p_T$  and (b) jet- $\eta$ . The  $b$ -jet trigger aware GRL has been applied.

#### 4.3.7 Efficiency Measurement and Systematic Derivation

In the previous two sections it has been shown that when applying a  $b$ -jet aware GRL, the  $b$ -jet trigger performance is understood and the data/simulation agreement is within 5%. In this section the measurement of data efficiency, data/simulation scale factors (SFs) and

associated systematics to account for the 5% are described.

As discussed above, there are two factors considered in this section. Firstly there is the  $\epsilon_{bTrig}$  measurement that accounts for differences in online and offline  $b$ -tagging given that a valid primary vertex has been found. Sections 4.3.7.1 to 4.3.7.3 describes the derivation of a set of systematics and corrections to the raw measurement and Section 4.3.7.4 presents the final measurement, which is applied as a jet-level efficiency in the final analysis. Secondly, in Section 4.3.7.5, is a description the measurement of the  $\epsilon_{bPerf}$  that accounts for the efficiency of finding a valid primary vertex and the relevant systematics, which is applied as an event level efficiency.

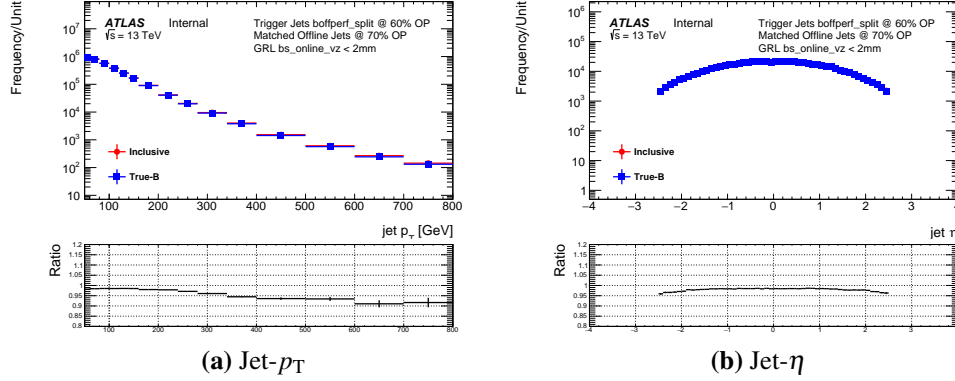
In this section describing the final measurement, the full 2016 data set is used, the simulated  $t\bar{t}$  sample includes single-top processes and the full event selection from Section 4.3.3 is applied.

#### 4.3.7.1 Purity Error

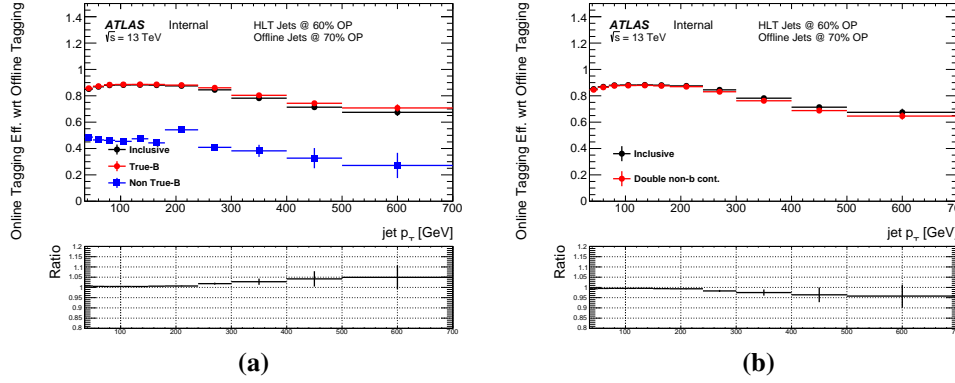
It is known that despite the strict event selection there will inevitably be non  $b$ -jet contamination in our sample. To estimate the  $b$ -jet purity simulation is used, where the true flavour of the jet is available. Jets are categorised as true  $b$ -jets, meaning that a  $B$ -hadron was found within a cone of  $R = 0.4$ , or true non- $b$ -jets if not. Then distributions for inclusive jets to the truth matched  $b$ -jets in the simulation sample are compared. Figure 4.12 shows the  $b$ -jet purity for jet- $p_T$  and jet- $\eta$ ; showing that the  $b$ -jet purity is  $> 95\%$  up to jet- $p_T \sim 300$  GeV and  $> 90\%$  for higher values of jet- $p_T$ .

To estimate the effect of these impurities on the efficiency measurement simulation is again used. Firstly, the efficiency in our nominal inclusive simulation is compared to the efficiencies if only true- $b$ -jets or true non- $b$ -jets are selected, this is shown in Figure 4.13(a). The ratio is applied as a correction to the final efficiency measurement. Then any mismodelling of the  $b$ -jet fraction in simulation is also considered, to account for this the efficiency for the simulated inclusive sample is compared to the efficiency when the non  $b$ -jet content has been doubled, as shown in Figure 4.13(b). The maximum difference from the efficiency measured in the inclusive simulated sample and the cases where there is only true  $b$ -jets and where the non  $b$ -jet content has been doubled, shown in the two ratio plots in Figure 4.13,

is taken as a symmetric systematic.



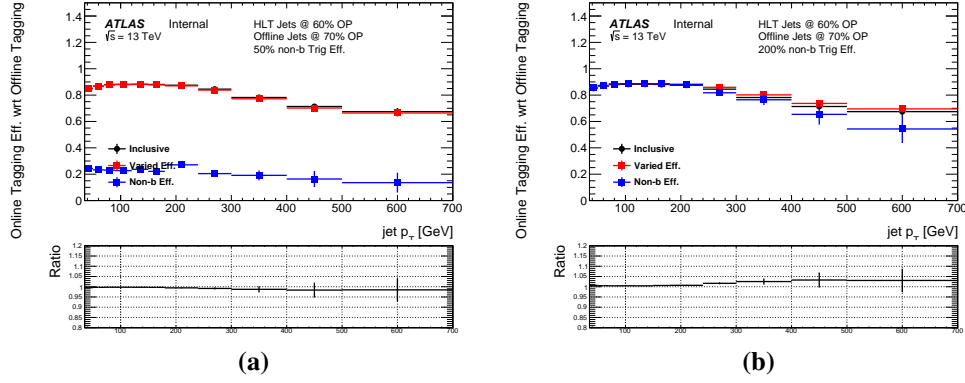
**Figure 4.12:** A comparison of offline jets tagged at the 70% operating point for inclusive jets (red) and truth-matched b-jets (blue) against  $\text{jet } p_T$  (a) and  $\text{jet } \eta$  (b) in a simulated  $t\bar{t}$  sample.



**Figure 4.13:** The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for inclusive jets (black) compared to truth matched b-jets and non b-jets (a) and the case where non b-jet content has been doubled (b) for a simulated  $t\bar{t}$  sample. The lower panel in both plots show the ratio to the inclusive efficiency.

#### 4.3.7.2 Non-b-jet trigger efficiency error

As one would expect and as shown in left plot of Figure 4.13, non b-jets (shown in blue) have a different  $b$ -jet trigger efficiency to that of b-jets. However the exact efficiency is not known well and could be mismodelled in simulation. To account for this uncertainty the nominal efficiency in simulation is compared to the cases where the non-b-jet efficiency has been halved and doubled in simulation, as shown in Figure 4.14. When doubling the non-b-jet trigger efficiency this is limited at the upper end to being no greater than the true b-jet trigger efficiency. The maximum bin-by-bin difference between the nominal and the two cases, as shown in the two ratio plots, is taken as a systematic.



**Figure 4.14:** The 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag for nominal inclusive case (black) compared to varied inclusive case (red) and just non  $b$ -jets (blue) in the case where non  $b$ -jet efficiency has been halved (a) and doubled (b) for a simulated  $t\bar{t}$  sample. The lower panel in both plots show the ratio of the varied inclusive efficiency to the nominal inclusive efficiency.

#### 4.3.7.3 High- $p_T$ extrapolation

Measuring  $b$ -jet trigger efficiency for high- $p_T$  jets is limited by the statistics in the simulated  $t\bar{t}$  sample, so the shape from simulation will be used to extrapolate the efficiency for  $\text{jet-}p_T > 240$  GeV. The point from which to extrapolate from was chosen as this is when data statistic error starts to become large.

The procedure is made of two sequential fits (normalisation and correction) to the data/simulation ratio, which are used to create a “corrected simulation”  $\epsilon_{bTrig}$  distribution. For  $\text{jet-}p_T > 240$  GeV, the corrected  $\epsilon_{bTrig}$  is used in place of data when measuring the data  $\epsilon_{bTrig}$  efficiency and when calculating data/MC scale factors. A final quadratic fit is used to assign a systematic.

In more detail:

- **Flat Normalisation Fit:**

The measured  $\epsilon_{bTrig}$ , in both data and simulation are compared, and a horizontal fit is performed to the ratio of the two. The fit range is set at  $p_T > 50$  GeV to discount the first bin, which has a larger purity uncertainty. This is then used to normalise the simulated efficiency distribution to match data. This fit is shown in the lower plot of panel (a) in Figure 4.15. The error on the one parameter of this fit is taken as a systematic error.

- **Linear Correction Fit:**

The measured  $\epsilon_{bTrig}$ , in both data and the normalised simulation are compared, and a linear fit is performed to the ratio of the two from  $\text{jet-}p_T > 240$  GeV. This is then used to correct the simulated efficiency distribution to match data. This fit is shown in the lower plot of panel (b) in Figure 4.15. The simulated  $\epsilon_{bTrig}$ , after both the normalisation and linear correction is referred to as the corrected simulation. To assign a systematic on the fit parameters, the slope of this fit is varied up and down within errors, whilst the point at which the fit crosses 1 is kept constant. The maximum difference between the nominal fit and the varied fits is taken as the error on the linear correction fit. Panel (c) of Figure 4.15 shows the data compared to the corrected simulation. The lower panel shows the ratio of the two, and the blue lines represent the errors on the linear correction fit.

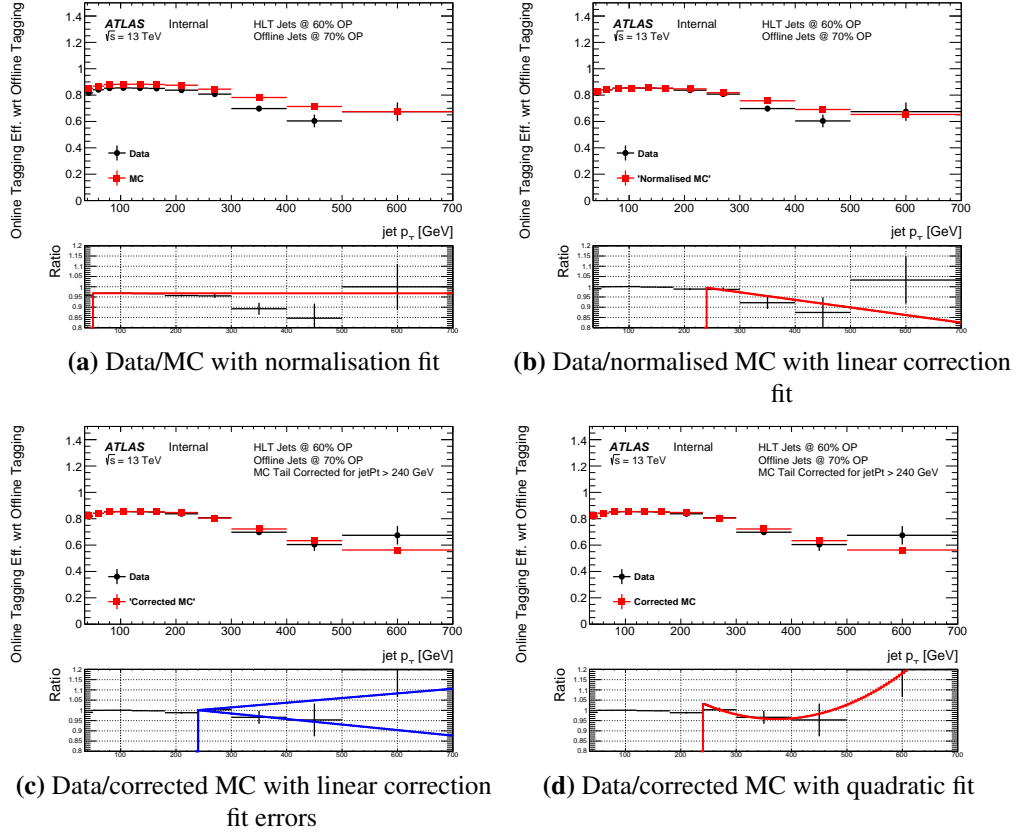
- **Quadratic Systematic Fit:**

Finally to assess an error on the choice of a linear fit as the functional form above, a fit is performed to the data and corrected simulation ratio using a quadratic function. This ratio and the fit is shown in panel (d) of Figure 4.15. The difference of the fit from 1 is considered as the functional form error when assigning as systematic.

The systematic error on the extrapolation is defined as the error from normalisation fit added to the bin-by-bin maximum of the error from the linear correction fit and the error from the quadratic systematic fit. The errors on the high- $p_T$  extrapolation procedure are summarised in Table 4.2

| Jet $p_T$ [GeV] | MC Extrap. Error (%) | Norm Fit Err. (%) | Lin. Fit (%) | Quad. Fit (%) |
|-----------------|----------------------|-------------------|--------------|---------------|
| 240.0-300.0     | 0.8                  | 0.0               | 0.8          | 0.3           |
| 300.0-400.0     | 4.0                  | 0.0               | 2.9          | 4.0           |
| 400.0-500.0     | 5.6                  | 0.0               | 5.6          | 1.7           |
| 500.0-700.0     | 18.0                 | 0.0               | 9.6          | 18.0          |

**Table 4.2:** A table showing the systematic assigned for the high- $p_T$  extrapolation.



**Figure 4.15:** A figure to demonstrate the high- $p_T$  extrapolation procedure for the 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag. Data (black) is compared against simulation (red) after various corrections have been applied as a function of jet- $p_T$ . Panel (a) shows the the flat normalisation fit uncorrected simulation, panel (b) shows the linear correction fit to normalised simulation, panel (c) shows the linear correction fit errors to the corrected simulation and panel (d) shows the quadratic fit to the corrected simulation.

#### 4.3.7.4 Jet-Level Efficiency and Scale Factor Measurement

Now the raw measurements of  $\epsilon_{bTrig}$  from Figure 4.10 and the additional corrections and systematics described above can be brought together. In Figure 4.10 it is shown that, whilst  $\epsilon_{bTrig}$  does depend on jet- $\eta$ , the data to simulation ratio is flat with respect to jet- $\eta$ . However there is no significant dependence on jet- $p_T$  hence data/simulation scale factors are derived as a function of only jet- $p_T$ .

The full jet-level  $\epsilon_{bTrig}$  measurement is shown in Figure 4.16. For use in combination with the simulation, a data/simulation scale factor as a function of jet- $p_T$  is also derived and will be applied at the jet-level, which is also shown in Figure 4.17.

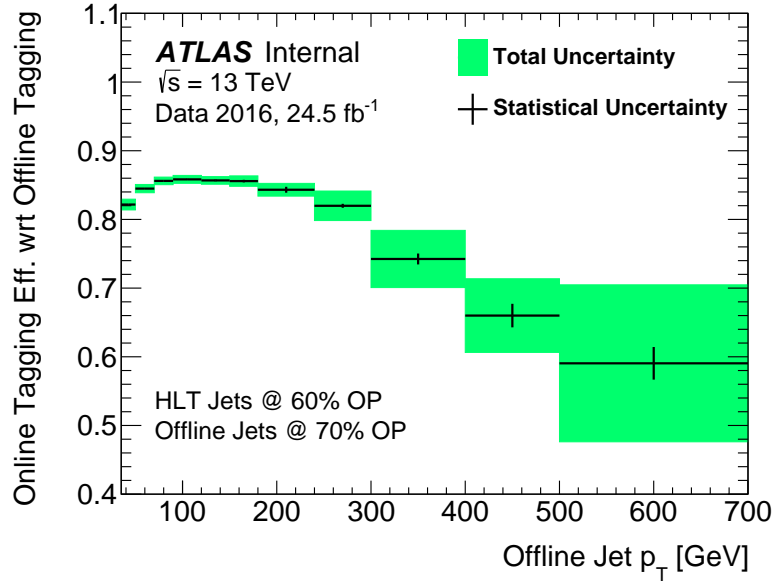


The errors considered for the jet-level efficiency account for: mismodelling of the  $b$ -jet purity in simulation, mismodelling of the  $b$ -jet trigger efficiency for non  $b$ -jets, simulation statistical error, data statistical error ( $\text{jet-}p_T < 240$  GeV) and simulation based extrapolation ( $\text{jet-}p_T > 240$  GeV). Table 4.3 summarises the errors on the jet-level scale factor. These errors are taken as a symmetric error in each  $\text{jet-}p_T$  bin and the scale factors are applied to each  $b$ -tagged jet.

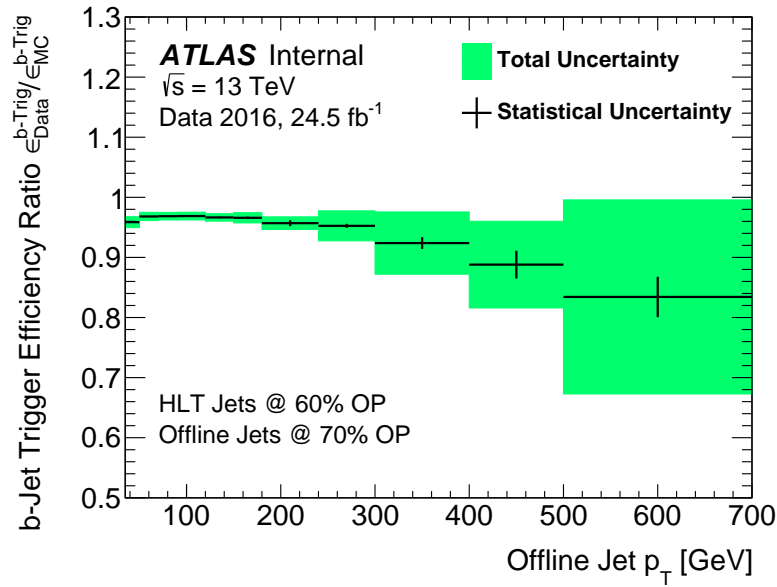
As a final sanity check Figure 4.18 shows  $\epsilon_{bTrig}$  measured in data to that from the corrected simulation, in the lower panel a ratio of data to corrected simulation is shown and the extrapolation and total errors are overlaid in red and green respectively. The derivation of the corrected simulation and associated extrapolation errors is described in Section 4.3.7.3 This shows that the corrected simulation lies within the total errors for the whole range of  $\text{jet-}p_T$  and at high- $p_T$ , as one might expect, the error is dominated by the extrapolation uncertainties. Note that the corrected simulation is only used to represent data for  $\text{jet-}p_T > 240$  GeV.

| Jet $p_T$ [GeV] | SF   | Total Err. (%) | Stat. (%) | Extrap. (%) | Pur. (%) | L. Trig. Eff. (%) |
|-----------------|------|----------------|-----------|-------------|----------|-------------------|
| 35.0-50.0       | 95.9 | 1.0            | 0.1       | -           | 0.7      | 0.7               |
| 50.0-70.0       | 96.8 | 0.7            | 0.1       | -           | 0.5      | 0.5               |
| 70.0-90.0       | 96.9 | 0.6            | 0.1       | -           | 0.5      | 0.5               |
| 90.0-120.0      | 96.9 | 0.7            | 0.1       | -           | 0.5      | 0.5               |
| 120.0-150.0     | 96.7 | 0.6            | 0.2       | -           | 0.4      | 0.4               |
| 150.0-180.0     | 96.6 | 0.9            | 0.2       | -           | 0.6      | 0.6               |
| 180.0-240.0     | 95.7 | 1.1            | 0.5       | -           | 0.7      | 0.7               |
| 240.0-300.0     | 95.3 | 2.6            | 0.4       | 0.8         | 1.8      | 1.7               |
| 300.0-400.0     | 92.4 | 5.6            | 1.1       | 4.0         | 2.8      | 2.5               |
| 400.0-500.0     | 88.8 | 8.1            | 2.6       | 5.6         | 4.2      | 3.3               |
| 500.0-700.0     | 83.4 | 19.4           | 4.0       | 18.0        | 4.9      | 3.1               |

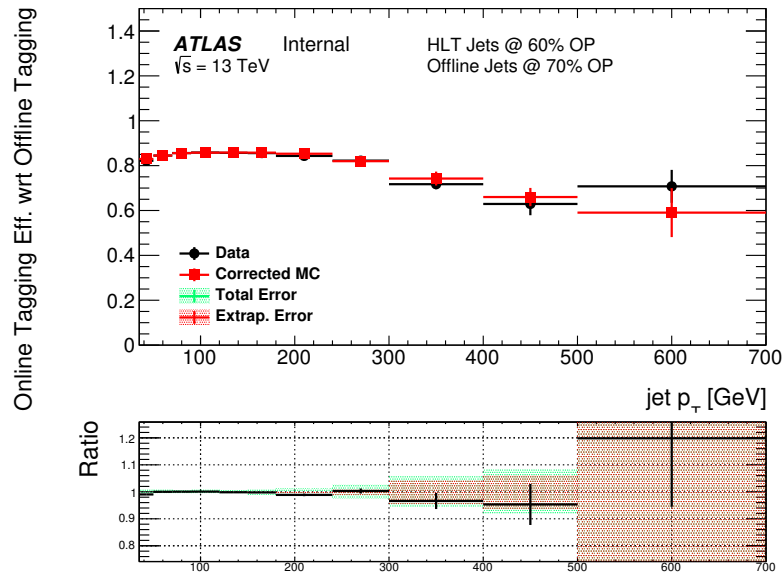
**Table 4.3:** A table showing the jet-level Data/simulation scale factor (SF) as a function of  $\text{jet-}p_T$  with total error and the contributions of the different systematics considered; specifically statistical, high- $p_T$  extrapolation, non- $b$ -jet purity and non- $b$ -jet trigger efficiency.



**Figure 4.16:** The measured 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag as measured in data as a function of offline jet- $p_T$ . The central values are shown in black with the statistical error and the green bands represent the total error including systematic errors.



**Figure 4.17:** Data/simulation scale factors for the 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag as a function of offline jet- $p_T$ . The central values are shown in black with the statistical error and the green bands represent the total error including systematic errors.



**Figure 4.18:** The measured 60%  $b$ -jet trigger efficiency with respect to an offline 70% operating point tag as measured in data (black) and from the corrected simulation (red) as a function of offline jet- $p_T$ . In the ratio plot on the lower panel the extrapolation errors is represented by the red band, whilst the total error is overlaid in green.

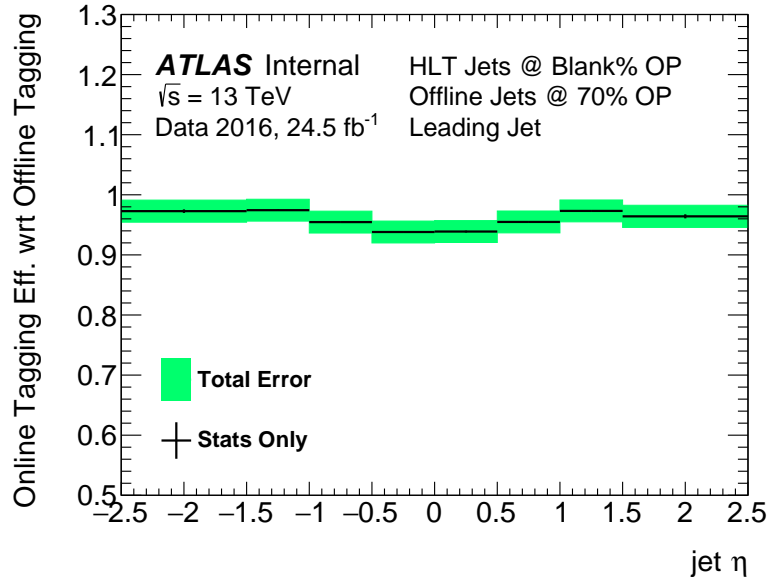
#### 4.3.7.5 Event-Level Efficiency and Systematic

As already discussed, in some regions of data-taking the performance *b*-jet trigger efficiency itself depends on the online beamspot position. Hence, a *b*-jet trigger aware GRL is applied to remove a large fraction of events where poor *b*-jet trigger performance is observed.

However, even after the application of this GRL, there remains a bias with respect to leading jet- $\eta$  in the probability of finding a valid primary vertex, which is notated as  $\epsilon_{bPerf}$ . This bias is shown in Figure 4.11. This efficiency is measured differently in each epoch, in Epoch 1 it can be found as the number of events with vertex class = 0 divided by the number of events, in Epoch 2 it is defined as the dividing the number of events that pass the trigger HLT\_mu26\_imedium\_2j35\_bperf by the number that pass the trigger HLT\_mu26\_imedium and in Epoch 3, due to the back-up vertex. It should be noted that this measurement made in each of the three regions separately and is then combined with each region weighted by its luminosity.

The value of  $\epsilon_{bPerf}$  is extremely close to 1 in simulation, in this case the efficiency in data and the scale factor are the same. To assign a systematic for this correction the statistical error in data and simulation in addition to a shape systematic are used. The shape systematic, to account for possible variations of the shape with respect to jet- $\eta$ , is defined as half of the difference between the maximum efficiency and the minimum efficiency in any jet- $\eta$  bin, which effectively covers a flat distribution with respect to jet- $\eta$  to one where the shape is twice as extreme as observed.

Table 4.4 and Figure 4.19 summarises the event-level efficiency correction and the associated systematics.



**Figure 4.19:** The measured  $\epsilon_{bPerf}$  as measured in data as a function of offline leading jet- $\eta$ . The central values are shown in black with the statistical error and the green bands represent the total error including systematic errors.

| Leading Jet $\eta$ | SF   | Total Error (%) | Data Stat. (%) | MC Stat. (%) | Shape Syst. (%) |
|--------------------|------|-----------------|----------------|--------------|-----------------|
| -2.5–1.5           | 97.3 | 1.9             | 0.3            | 0.1          | 1.9             |
| -1.5–1.0           | 97.4 | 1.9             | 0.1            | 0.0          | 1.9             |
| -1.0–0.5           | 95.5 | 1.9             | 0.1            | 0.0          | 1.9             |
| -0.5–0.0           | 93.8 | 1.9             | 0.2            | 0.0          | 1.9             |
| 0.0–0.5            | 93.9 | 1.9             | 0.2            | 0.0          | 1.9             |
| 0.5–1.0            | 95.5 | 1.9             | 0.2            | 0.0          | 1.9             |
| 1.0–1.5            | 97.3 | 1.9             | 0.1            | 0.0          | 1.9             |
| 1.5–2.5            | 96.4 | 1.9             | 0.3            | 0.1          | 1.9             |

**Table 4.4:** A table showing the event-level Data/MC scale factor (SF) as a function of leading jet- $\eta$  with total error and the contributions of the different systematics considered.

### 4.3.8 Cross-checks

#### 4.3.8.1 Simulation checks

- Ttbar alone vs ttbar+tW
- Try powheg

#### 4.3.8.2 Electron/Muon overlap checks

#### 4.3.8.3 Event Level Eff: Showing correlation with $z_{bs}^{online}$

- Show that it comes from high beamspot z-position only.
- i.e.  $\epsilon_{bPerf}$  vs eta for different bs regions.

#### 4.3.8.4 Event Level Eff: Re-weighting of sub-leading jet

- We did a test where we applied correction to leading and shows the subleading was flat within systematics (2%)

Any others that are good?

Cross-checks can be moved to appendix

### 4.4 To Do

These can be considered on my list. - Cite in plot caption

- Uncertainty instead of error
- Update plots to most current version (and label those that are not)
- In caption I want (a) before plot i.e. (a) jet-pT, (b) jet-eta. - Always use data/simulation instead of data/MC
- use Epoch instead of epoch

## Chapter 5

# General Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

## **Appendix A**

# **An Appendix About Stuff**

(stuff)



## **Appendix B**

# **Another Appendix About Things**

(things)

## Appendix C

# Colophon

*This is a description of the tools you used to make your thesis. It helps people make future documents, reminds you, and looks good.*

(example) This document was set in the Times Roman typeface using L<sup>A</sup>T<sub>E</sub>X and BibT<sub>E</sub>X, composed with a text editor.

# Bibliography

- [1] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at  $s^{1/2} = 540$  GeV. *Phys. Lett.*, 122B:103–116, 1983. [611(1983)].
- [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.*, 126B:398–410, 1983.
- [3] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett.*, 122B:476–485, 1983.
- [4] P. Bagnaia et al. Evidence for  $Z_0 \rightarrow e^+e^-$  at the CERN anti-p p Collider. *Phys. Lett.*, 129B:130–140, 1983.
- [5] F. Abe et al. Observation of top quark production in  $\bar{p}p$  collisions with the collider detector at fermilab. *Phys. Rev. Lett.*, 74:2626–2631, Apr 1995.
- [6] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, Apr 1995.
- [7] Atlas public lumi results run-2. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [8] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [9] A. Airapetian et al. *ATLAS detector and physics performance: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1999.

- [10] G. Aad et al. Performance of the atlas detector using first collision data. *Journal of High Energy Physics*, 2010(9):56, Sep 2010.
- [11] C Gutsche. First observation of electroweak z boson plus two jet production, October 2014.
- [12] M Capeans, G Darbo, K Einsweiler, 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.
- [13] C Gemme. The ATLAS Insertable B-Layer (IBL) Project. Jul 2011.
- [14] Claude Leroy and Pier-Giorgio Rancoita. *Principles of Radiation Interaction in Matter and Detection*. World Scientific, 2016.
- [15] Katherine Pachal. Search for new physics in the dijet invariant mass spectrum at 8 TeV, Jan 2015. Presented 01 Jun 2015.
- [16] Michele Livan and Richard Wigmans. Misconceptions about Calorimetry. *Instruments*, 1(1):3, 2017.
- [17] Lene Bryngemark, Torsten AKESSON, Else LYTKEN, and Johan RATHSMAN. Search for new phenomena in dijet angular distributions at  $s = 8$  and 13 TeV, Feb 2016. Presented 18 Mar 2016.
- [18] J.J. Goodson. *Search for Supersymmetry in States with Large Missing Transverse Momentum and Three Leptons including a Z-Boson*. PhD thesis, Stony Brook University, May 2012. Presented 17 Apr 2012.
- [19] Georges Aad et al. Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System. Technical Report CERN-LHCC-2013-018. ATLAS-TDR-023, Sep 2013. Final version presented to December 2013 LHCC.
- [20] Morad Aaboud et al. Performance of the ATLAS Trigger System in 2015. *Eur. Phys. J.*, C77(5):317, 2017.
- [21] H. Kirschenmann. Jets at cms and the determination of their energy scale. <http://cms.web.cern.ch/news/jets-cms-and-determination-their-energy-scale>.

- [22] G. Aad and etal. Topological cell clustering in the atlas calorimeters and its performance in lhc run 1. *The European Physical Journal C*, 77(7):490, Jul 2017.
- [23] Yu Nakahama. The atlas trigger system: Ready for run-2. *Journal of Physics: Conference Series*, 664(8):082037, 2015.
- [24] M. Aaboud et al. Search for pair production of higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton-proton collisions at  $\sqrt{s} = 13$  TeV with the atlas detector. *Phys. Rev. D*, 94:052002, Sep 2016.
- [25] 2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data. Technical Report ATL-DAQ-PUB-2016-001, CERN, Geneva, Mar 2016.
- [26] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.
- [27] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.
- [28] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.
- [29] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev. D*, 82:074018, 2010.
- [30] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152, 2001.