Reconstructing a Quark and Gluon Jet Response at ATLAS

by

James Walter Beare

B.Sc., University of Waterloo, 2013

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in the Department of Physics Faculty of Science

© James Walter Beare 2016
SIMON FRASER UNIVERSITY Fall 2016

All rights reserved. However, in accordance with the Copyright Act of Canada, this work may be reproduced without authorization under the conditions for “Fair Dealing.” Therefore, limited reproduction of this work for the purposes of private study, research, education, satire, parody, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
Approval

Name: James Walter Beare
Degree: Master of Science (Physics)
Title: Reconstructing a Quark and Gluon Jet Response at ATLAS

Examining Committee: Chair: Eldon Emberly
Professor
Michel Vetterli
Senior Supervisor
Professor

Bernd Stelzer
Supervisor
Associate Professor

Andrew DeBenedictis
Supervisor
Senior Lecturer

Sarah Johnson
Internal Examiner
Senior Lecturer

Date Defended: 2 September 2016
Abstract

Jets, collimated sprays of particles, are the most commonly produced objects in high energy subatomic collisions. Jets represent the final states of quarks and gluons produced during collisions. The fraction of a jet’s energy that is measured by a calorimeter is called the response. Quark jets (jets initiated by quarks) and gluon jets (jets initiated by gluons) have a different response in the calorimeter. In this thesis the response of quark and gluon jets is reconstructed using jets in dijet and photon + jet events. To measure jet response in dijet events a method is developed to correct the energy of one of the jets so that it may be used as a reference object in the calibration procedure. The reconstructed dijet, quark and gluon responses are shown to agree with Monte Carlo simulation predictions within their uncertainties.

Keywords: Jets, Quarks, Gluons, Jet Response, MPF
Dedication

I would like to dedicate this thesis to my family.
Acknowledgements

I would like to take this opportunity to thank everyone who has helped me over the last three years with this project as well as my life away from it. A special thank you to both Mike and Jamie for their guidance in the project.
Contents

Approval ii
Abstract iii
Dedication iv
Acknowledgements v
List of Tables ix
List of Figures x

1 Introduction 1
   1.1 The Standard Model ........................................... 2
   1.2 Units and Conventions ........................................ 3

2 Calorimetry 5
   2.1 Electromagnetic Calorimetry ................................. 6
      2.1.1 Photon Interactions ..................................... 6
      2.1.2 Electrons and Positrons ................................ 7
      2.1.3 Electromagnetic Showers ................................. 9
      2.1.4 Response to Electromagnetic Showers .................... 12
   2.2 Hadronic Calorimetry ........................................ 13
      2.2.1 Electromagnetic Shower Component ....................... 13
      2.2.2 Nuclear Interactions .................................... 13
      2.2.3 Hadronic Showers ....................................... 14
      2.2.4 Response to Hadronic Showers .......................... 14

3 Experimental Setup 17
   3.1 The Large Hadron Collider ................................ 17
   3.2 The ATLAS Detector ...................................... 18
      3.2.1 The ATLAS Coordinate System ........................ 18
      3.2.2 ATLAS Inner Detector ............................... 20
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.3</td>
<td>ATLAS Calorimeters</td>
<td>21</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Muon Spectrometer</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>Physics Object Reconstruction</td>
<td>30</td>
</tr>
<tr>
<td>4.1</td>
<td>Calorimeter Clusters</td>
<td>30</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Sliding Window Clusters</td>
<td>30</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Topological Clusters</td>
<td>31</td>
</tr>
<tr>
<td>4.2</td>
<td>Jets</td>
<td>31</td>
</tr>
<tr>
<td>4.3</td>
<td>Muons</td>
<td>33</td>
</tr>
<tr>
<td>4.4</td>
<td>Missing Tranverse Energy</td>
<td>33</td>
</tr>
<tr>
<td>4.5</td>
<td>Triggers</td>
<td>34</td>
</tr>
<tr>
<td>5</td>
<td>Jet Energy Scale</td>
<td>35</td>
</tr>
<tr>
<td>5.1</td>
<td>ATLAS JES Calibration Framework</td>
<td>35</td>
</tr>
<tr>
<td>5.2</td>
<td>Jet Response</td>
<td>38</td>
</tr>
<tr>
<td>5.3</td>
<td>In-situ Techniques</td>
<td>39</td>
</tr>
<tr>
<td>5.4</td>
<td>Quark and Gluon Jets</td>
<td>41</td>
</tr>
<tr>
<td>6</td>
<td>Jet Properties and Response Correction</td>
<td>44</td>
</tr>
<tr>
<td>6.1</td>
<td>Jet Variable Study</td>
<td>44</td>
</tr>
<tr>
<td>6.2</td>
<td>Correcting the Reference Jet Energy</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>Determining the Jet Response</td>
<td>67</td>
</tr>
<tr>
<td>7.1</td>
<td>Data Samples</td>
<td>67</td>
</tr>
<tr>
<td>7.2</td>
<td>Selection Criteria</td>
<td>67</td>
</tr>
<tr>
<td>7.2.1</td>
<td>photon selection</td>
<td>67</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Jet Selection</td>
<td>68</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Event Selection</td>
<td>69</td>
</tr>
<tr>
<td>7.3</td>
<td>Measuring the Jet Response</td>
<td>72</td>
</tr>
<tr>
<td>7.4</td>
<td>Systematic Uncertainties</td>
<td>72</td>
</tr>
<tr>
<td>7.4.1</td>
<td>ISR/FSR</td>
<td>73</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Signal/Background Definitions</td>
<td>73</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Correlation Effects</td>
<td>76</td>
</tr>
<tr>
<td>7.4.4</td>
<td>Alpha Closure σ</td>
<td>77</td>
</tr>
<tr>
<td>7.4.5</td>
<td>Photon + Jet Systematic Uncertainty</td>
<td>77</td>
</tr>
<tr>
<td>7.5</td>
<td>Results</td>
<td>79</td>
</tr>
<tr>
<td>8</td>
<td>Calculating the Quark and Gluon Jet Responses</td>
<td>82</td>
</tr>
<tr>
<td>8.1</td>
<td>Calculating Uncertainties</td>
<td>82</td>
</tr>
<tr>
<td>8.2</td>
<td>Results</td>
<td>85</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1  Average fractional energy deposition of the non-electromagnetic com-
ponent of a 1.3 GeV pion along with the invisible energy [1]. . . . . . . 14

Table 3.1  The angular coverage and number of channels in the ATLAS ID . . . 21

Table 4.1  Window dimensions ($\eta \times \phi$) for electrons, converted photons and un-
converted photons [11] . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

Table 5.1  Colour factor for single gluon exchange interactions [16] . . . . . . . 42

Table 6.1  Jet variables and a brief description . . . . . . . . . . . . . . . . . . . 47

Table 7.1  Photon Triggers . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 68

Table 7.2  Jet Triggers for calibrated scale $p_T^{avg}$ . . . . . . . . . . . . . . . . 69
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Standard Model Particles</td>
<td>3</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Photon Interactions as a function of $Z$ and energy</td>
<td>7</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Photon Interaction Cross-sections</td>
<td>8</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Fractional energy loss in lead</td>
<td>10</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Simple electromagnetic shower</td>
<td>11</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Electron and photon resolution at ATLAS</td>
<td>12</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>LHC with Injectors</td>
<td>18</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>An overview of the ATLAS detector showing the various subdetectors [9]</td>
<td>19</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>The ATLAS Inner Detector [9]</td>
<td>22</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>A charged particle traverses the ATLAS Inner Detector [9]</td>
<td>23</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>The ATLAS Calorimeter [9]</td>
<td>25</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Accordion structure of the Electromagnetic Barrel Calorimeter [9]</td>
<td>26</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>Calorimeter thickness as a function of pseudorapidity [9]</td>
<td>27</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>Schematic of a section of the TILE calorimeter [9]</td>
<td>28</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>The Muon spectrometer subdetectors [9]</td>
<td>29</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>2011 quark and gluon response difference</td>
<td>35</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Jet Energy Scale Uncertainties</td>
<td>36</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>ATLAS Jet Construction and Calibration Scheme</td>
<td>37</td>
</tr>
<tr>
<td>Figure 5.4</td>
<td>Jet Vertex Fraction</td>
<td>39</td>
</tr>
<tr>
<td>Figure 5.5</td>
<td>Quark versus Gluon Jets</td>
<td>42</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>Monte Carlo Response</td>
<td>44</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>The cluster and string fragmentation models</td>
<td>46</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>Variable distributions in data and simulation</td>
<td>48</td>
</tr>
<tr>
<td>Figure 6.4</td>
<td>NTrk500 signal and background distributions</td>
<td>49</td>
</tr>
<tr>
<td>Figure 6.5</td>
<td>TrackPt500Frac signal and background distributions</td>
<td>49</td>
</tr>
<tr>
<td>Figure 6.6</td>
<td>AVGTrackPt1Frac signal and background distributions</td>
<td>50</td>
</tr>
<tr>
<td>Figure 6.7</td>
<td>TrackWidth signal and background distributions</td>
<td>50</td>
</tr>
<tr>
<td>Figure 6.8</td>
<td>Width signal and background distributions</td>
<td>51</td>
</tr>
<tr>
<td>Figure 6.9</td>
<td>EMBFrac signal and background distributions</td>
<td>51</td>
</tr>
<tr>
<td>Figure 6.10</td>
<td>EMB1Frac signal and background distributions</td>
<td>52</td>
</tr>
<tr>
<td>Figure 6.11</td>
<td>EMB3Frac signal and background distributions</td>
<td>52</td>
</tr>
<tr>
<td>Figure 6.12</td>
<td>Mass signal and background distributions</td>
<td>53</td>
</tr>
<tr>
<td>Figure 6.13</td>
<td>TILEB0Frac signal and background distributions</td>
<td>54</td>
</tr>
<tr>
<td>Figure 6.14</td>
<td>EMB2OverEMB signal and background distributions</td>
<td>54</td>
</tr>
<tr>
<td>Figure 6.15</td>
<td>TILEB1OverTILE signal and background distributions</td>
<td>55</td>
</tr>
<tr>
<td>Figure 6.16</td>
<td>MostELayer signal and background distributions</td>
<td>56</td>
</tr>
<tr>
<td>Figure 6.17</td>
<td>MostEperX0 signal and background distributions</td>
<td>56</td>
</tr>
<tr>
<td>Figure 6.18</td>
<td>EndLayer signal and background distributions</td>
<td>57</td>
</tr>
<tr>
<td>Figure 6.19</td>
<td>NumTowers signal and background distributions</td>
<td>57</td>
</tr>
<tr>
<td>Figure 6.20</td>
<td>Probability distribution for EMBFrac</td>
<td>58</td>
</tr>
<tr>
<td>Figure 6.21</td>
<td>Signal-weighted probability distribution for EMBFrac</td>
<td>59</td>
</tr>
<tr>
<td>Figure 6.22</td>
<td>Event-weighted signal/background distribution</td>
<td>59</td>
</tr>
<tr>
<td>Figure 6.23</td>
<td>Correlations between jet variables in Pythia for reconstructed $p_T$ from 105-125 GeV</td>
<td>61</td>
</tr>
<tr>
<td>Figure 6.24</td>
<td>Correlations between jet variables in Pythia for reconstructed $p_T$ from 500-600 GeV</td>
<td>62</td>
</tr>
<tr>
<td>Figure 6.25</td>
<td>MCResponse vs $-\log(L)$ in the momentum range 105-125 GeV</td>
<td>63</td>
</tr>
<tr>
<td>Figure 6.26</td>
<td>Fitted response distributions when binning in $-\log(L)$</td>
<td>64</td>
</tr>
<tr>
<td>Figure 6.27</td>
<td>The mean response fit vs likelihood</td>
<td>65</td>
</tr>
<tr>
<td>Figure 6.28</td>
<td>The alpha closure as a function of momentum in Pythia</td>
<td>65</td>
</tr>
<tr>
<td>Figure 6.29</td>
<td>The standard deviation of the alpha closure as a function of momentum in Pythia</td>
<td>66</td>
</tr>
<tr>
<td>Figure 7.1</td>
<td>The leading Jet $p_T$ in dijet events</td>
<td>70</td>
</tr>
<tr>
<td>Figure 7.2</td>
<td>The subleading jet $p_T$ in dijet events</td>
<td>70</td>
</tr>
<tr>
<td>Figure 7.3</td>
<td>The fraction of leading jet $p_T$ carried by the third-leading jet in dijet events</td>
<td>71</td>
</tr>
<tr>
<td>Figure 7.4</td>
<td>$\Delta\phi$ between the leading and subleading</td>
<td>71</td>
</tr>
<tr>
<td>Figure 7.5</td>
<td>The dijet response distribution</td>
<td>72</td>
</tr>
<tr>
<td>Figure 7.6</td>
<td>Relative response difference between the nominal and varied ISR/FSR cuts in dijet data events</td>
<td>74</td>
</tr>
<tr>
<td>Figure 7.7</td>
<td>Relative response difference between the nominal and varied ISR/FSR cuts in dijet PYTHIA events</td>
<td>74</td>
</tr>
<tr>
<td>Figure 7.8</td>
<td>Relative response difference between the nominal and varied ISR/FSR cuts in $\gamma +$ jet Data events</td>
<td>75</td>
</tr>
<tr>
<td>Figure 7.9</td>
<td>Relative response difference between the nominal and varied ISR/FSR cuts in Pythia $\gamma + \text{jet}$ events</td>
<td>75</td>
</tr>
<tr>
<td>Figure 7.10</td>
<td>Relative response difference between the nominal and varied signal-background definitions in data dijet events</td>
<td>76</td>
</tr>
<tr>
<td>Figure 7.11</td>
<td>Relative response difference between the nominal and varied signal-background definitions in Pythia dijet events</td>
<td>77</td>
</tr>
<tr>
<td>Figure 7.12</td>
<td>The relative response differences when using the nominal 14 variable log-likelihood functions and the varied 6 and 9 variable log-likelihood functions in data dijet events</td>
<td>78</td>
</tr>
<tr>
<td>Figure 7.13</td>
<td>The relative response differences when using the nominal 14 variable log-likelihood functions and the varied 6 and 9 variable log-likelihood functions in Pythia events</td>
<td>78</td>
</tr>
<tr>
<td>Figure 7.14</td>
<td>Dijet MPF Response</td>
<td>80</td>
</tr>
<tr>
<td>Figure 7.15</td>
<td>$\gamma + \text{jet}$ MPF Response</td>
<td>81</td>
</tr>
<tr>
<td>Figure 8.1</td>
<td>Parton fractions in dijet events</td>
<td>83</td>
</tr>
<tr>
<td>Figure 8.2</td>
<td>Parton fraction in $\gamma + \text{jet}$ events</td>
<td>83</td>
</tr>
<tr>
<td>Figure 8.3</td>
<td>The $c$ quark jet response</td>
<td>84</td>
</tr>
<tr>
<td>Figure 8.4</td>
<td>Quark Jet Response</td>
<td>86</td>
</tr>
<tr>
<td>Figure 8.5</td>
<td>The Gluon jet response</td>
<td>87</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Jets are the most common objects produced in proton-proton colliders like the Large Hadron Collider (LHC). A jet is the final state of quarks and gluons produced during collisions. Quarks and gluons produce many particles which are eventually detected in a calorimeter. These particles are the constituents of a jet. The jet energy is the total energy of its constituents and can be measured when the constituents deposit energy in the calorimeter, which is described in Chapter 2.

The measured energy of a jet is always lower than its true energy because many of the jet constituents interact through nuclear interactions. Energy is lost when an atomic nucleus breaks apart or decays from an excited state. The fraction of the jet energy that is measured by the calorimeter compared to its true energy is known as the response of the calorimeter to the jet. This is often referred to by the short-hand expression ‘jet response’.

The ATLAS collaboration calibrates the response of jets by using momentum balance and compares the jet momentum to the momentum of a well measured reference object like a photon. One of the techniques used to do this is the Missing transverse energy Project Fraction (MPF). Quark jets are jets which were initiated by light-quarks (u,d,s) and gluon jets are jets which were initiated by gluons. Photon + jet events are dominated by quark jets, meaning the measured calibration applies mostly to quark jets. Applying the derived calibration to samples with a large amount of jets initiated by gluons requires additional uncertainties to be added. The goal of this thesis is to reconstruct a quark and gluon jet response to reduce the jet momentum uncertainty in samples which have a large fraction of gluon jets.

To calculate a quark and gluon response requires that the response of two samples with a sufficiently different quark and gluon jet fraction be measured. The two samples chosen for the research in this thesis are two-jet (dijet) and photon + jet events. These two samples have a relatively large difference in their quark and gluon jet fractions until high energy is reached. In this thesis a response correction is developed based on jet properties that allows a jet in a dijet event to be used as a reference object in the MPF technique.
By measuring a quark and gluon jet response, the response of jets in any sample could be calculated using the formula,

\[ R_S = R^q \times f_S^q + R^g \times f_S^g, \]  

(1.1)
where \( R_S \) is the response of the sample, \( R^q \) is the response of quark jets, \( R^g \) is the response of gluon jets and \( f_S^q \) and \( f_S^g \) are the fractions of quark and gluon jets in the sample.

### 1.1 The Standard Model

The Standard Model of Particle Physics (SM) is a theoretical framework which describes physics at the subatomic level. The SM is a relativistic quantum field theory acting on particles which are point-like, meaning they have no physical size. In the framework, matter consists of half-integer spin particles called fermions with the forces between them mediated by integer-spin particles called bosons. The bosons are often referred to as gauge bosons because they are associated with a gauge symmetry of the force which they carry.

The Standard Model consists of the Glashow-Salam-Weinberg model (GSW) along with Quantum Chromodynamics (QCD). The GSW model describes the electro-weak force which affects all electrically charged and weakly interacting particles. GSW describes the unification of the weak nuclear and electromagnetic forces. In the theory the Z, W± and photon mediate interactions between electrically charged and weakly interacting particles.

QCD describes the strong force which affects quarks and gluons collectively known as partons. Quarks are spin-\( \frac{1}{2} \) and the strong force is mediated by spin-1 gluons. In the theory quarks and gluons carry color charge analogous to the electric charge of electromagnetism. In QCD there are three possible color charges labelled red, green and blue. The theory is complicated by the fact that gluons only exist in colored states governed by an SU(3) group so that they experience the strong force themselves. Gluons are electrically neutral so they do not experience the electromagnetic force. Quarks on the other hand are electrically charged so they do experience the electromagnetic force.

The strong force gets stronger with increasing distance which plays a part in the phenomenon of colour confinement. Colour confinement means that coloured objects cannot be observed on their own and must exist in colour neutral combinations. These colour neutral combinations form stable particles called hadrons which are held together by the strong force. The type of hadron is determined by the valence quarks inside the hadron. Quark-antiquark hadrons are known as mesons while three quark hadrons are known as baryons.

Particles which do not experience the strong force are known as leptons. There are six flavours of quarks and six flavours of leptons. The quark flavours are up, down, strange, charm, top and bottom. Lepton flavours include the electron, muon, and tau along with
their corresponding neutrinos. Both quarks and leptons are divided into three generations by mass. Each quark generation consists of a positively and a negatively electrically charged partner. Figure 1.1 summarizes the properties of the various quarks, leptons and gauge bosons.

![Figure 1.1: The Standard Model particles](image)

The last piece of the Standard Model is the Higgs boson. In the electroweak theory the masses of the Z and $W^\pm$ bosons were predicted to be zero, strongly conflicting with their observed masses given in Fig. 1.1. A way to give these particles mass without destroying gauge invariance was needed. This was resolved by adding a scalar field which spontaneously breaks the symmetry of the electroweak ground state. This scalar field is also used to explain how other particles gain mass. As a consequence of this new field, the Higgs boson was predicted to exist. In 2012 both ATLAS and CMS announced the discovery of the Higgs boson at around 125 GeV.

1.2 Units and Conventions

In this thesis the speed of light, $c$, and the reduced Plank constant, $\hbar$, are taken to be 1 unless specifically stated or expressed in a formula. The charge of a particle is given in multiples of the electron charge. Energy is measured in electron Volts (eV) which is the amount of energy gained by a particle with unit charge across a potential difference of 1 Volt. One eV is equal to $1.602 \times 10^{-19}$ Joules. In this thesis Mega-electron Volts (MeV,
$10^6 \text{ eV}$, Giga-electron Volts (GeV, $10^9 \text{ eV}$) and Tera-electron Volts (TeV, $10^{12} \text{ eV}$) are most often used.

The probability of a scattering event is given by the cross-section, $\sigma$, typically measured in barns (b). 1 b is equal to $10^{-28} \text{ m}^2$. Data are typically quantified by the integrated luminosity, $L$, which is measured in inverse barns. The number of events of a particular type in a dataset can be determined by multiplying the integrated luminosity by the cross-section of that type of event. This thesis typically uses femtobarns, (fb, $10^{-15} \text{ b}$) to describe cross sections and inverse-femtobarns to describe luminosities.
Chapter 2

Calorimetry

A calorimeter measures the energy of a particle by fully absorbing its energy through the process of the particle shower it generates. A shower is a stream of particles created through interactions of high-energy particles with calorimeter material. A sampling calorimeter layers active material, designed to measure the energy of showers, with an absorber material, designed to efficiently create showers. The energy deposited in the active layer compared to the total energy deposited in the calorimeter is known as the sampling fraction and is an important characteristic of calorimeters. Another type of calorimeter is a homogeneous calorimeter made entirely of active material. Homogeneous calorimeters tend to have better energy resolution than sampling calorimeters but they require a much larger space and are much more prone to radiation damage. In both cases the measurement is a destructive process since the original particle is absorbed. For this reason calorimeters must be located outside other subdetectors such as tracking chambers.

Calorimeters are generally optimized for either electromagnetically or hadronically interacting particles and are referred to as Electromagnetic and Hadronic calorimeters respectively. Electromagnetic calorimeters are designed to measure accurately the energy of electrons, positrons and photons. This type of calorimeter will also absorb energy from any charged particle but higher mass particles such as charged hadrons can pass through the electromagnetic calorimeter while only depositing as little as 20-30% of their energy. ATLAS uses sampling calorimeters with Liquid Argon (LAr) active layers and lead for the passive material to measure electromagnetically interacting particles. ATLAS also makes use of hadronic calorimeters to complete the energy measurement of hadrons which pass through the electromagnetic calorimeters. In the barrel (end cap) region ATLAS uses plastic scintillator (LAr) as the active material and steel (copper) as the absorber material. The most forward regions have their own dedicated calorimeters designed to deal with the extra radiation these regions receive. The calorimeter in this region layers one level of electromagnetic calorimeter (LAr and copper) with two layers of hadronic calorimeters (LAr and tungsten).
Section 2.1 gives an introduction to electromagnetic calorimeters. The section begins with a description of electromagnetic interactions and goes on to describe electromagnetic showers. In section 2.2 hadronic calorimetry and hadronic showers are discussed.

2.1 Electromagnetic Calorimetry

Electromagnetic calorimeters accurately measure the energy of photons, electrons and positrons. At different energies separate processes dominate the interaction of both photons and electrons. The general structure of these processes involves a photon or electron interacting with the Coulomb field of atoms within the detector. This interaction depends on the atomic number $Z$ of the material. While these processes are electromagnetic in nature, they are different for photons and electrons.

2.1.1 Photon Interactions

Photons interact with matter in four different ways: the photoelectric effect, Rayleigh scattering, Compton scattering, and pair production. At low energy (below a few hundred keV for most materials) the photoelectric effect is the dominant interaction between photons and atoms. In this process a photon is absorbed by an atom and an electron is emitted. The excited atom returns to the ground state through the emission of (lower energy) photons or Auger electrons.\footnote{Classically, an Auger electron is caused by an electron falling from a higher energy level to a lower energy level, releasing a photon. This photon is absorbed by a valence electron which is ejected from the atom. Quantum mechanically this is one process.} The photoelectric cross section decreases with photon energy as $E^{-3}$ and has a dependence on $Z$ between $Z^4$ and $Z^5$ [1].

At low energies Rayleigh scattering is also important. In this process atomic electrons cause the photon to be deflected in a new direction. Rayleigh scattering does not contribute to the energy deposition of the photon as the energy of the photon remains the same. This process does however change the spatial distribution of the energy deposition.

As the energy of the photon increases above a few hundred keV, Compton scattering becomes the dominant interaction between photons and atoms. In Compton scattering a photon is scattered off of an atomic electron, ejecting the electron from the atom and reducing the energy of the photon. The electron and photon are scattered in different directions with respect to the original direction of the photon. The cross section for this process falls as $E^{-1}$ and has a maximum when the scattered electron and photon are coincident with the incoming photon. The Compton scattering cross section is approximately proportional to $Z$. For most materials used in calorimeters this effect is dominant up to about 5 MeV, at which point pair production becomes important [1].

At energies above twice the electron mass (511 keV) pair production becomes possible. In pair production a photon in the presence of a charged particle may create an
Figure 2.1: The dominant photon interactions are shown as a function energy and Z. The red line shows where Compton scattering becomes more dominant than the photoelectric effect and the green line shows where pair production becomes more important than Compton scattering [30].

Electrons and positrons lose energy primarily through bremsstrahlung and ionisation. Bremsstrahlung is radiation produced by accelerating (or decelerating) charged particles in the presence of electromagnetic fields\(^2\). In 1 cm of lead a multi-GeV electron will emit thousands of photons in the form of bremsstrahlung. Most of the photons are low energy, in the eV to MeV range but some can be in the GeV range [1]. The energy loss of electrons through bremsstrahlung is given by,

\[^2\text{Usually the electromagnetic field of a nucleus but this can also be the electromagnetic field of atomic electrons.}\]
Figure 2.2: The cross-section of different photon interactions in lead as a function of energy. The photoelectric cross section, \( \sigma_{P.E.} \), dominates at low energy with contributions from Rayleigh scattering. In the MeV range Compton scattering dominates while at high energy pair production in the presence of nuclear electromagnetic fields, \( \kappa_{nuc} \) dominates [2].

\[
\frac{dE}{dx} = n_e E \sigma_{brem},
\]

(2.1)

where \( \sigma_{brem} \), the bremsstrahlung cross-section, is approximately

\[
\sigma_{brem} = 4Z^2r_e^2\alpha(\ln \frac{184}{Z} + \frac{1}{18})
\]

(2.2)

where \( r_e \) is the classical electron radius and \( \alpha \) is the fine structure constant [3, 4].

Photons from bremsstrahlung can eject atomic electrons and ionise the atom. If energetic enough, the ejected electron can also ionise atoms through the same process. The energy loss of an electron due to ionisation per length \( dx \) is given by,

\[
\frac{dE}{dx} = \frac{2\pi n_e e^4}{m_e c^2} (2\ln \frac{2m_e c^2}{I} + 3\ln \gamma - 2),
\]

(2.3)

where \( n_e \) is the electron density in the medium, \( I \) is the mean ionization potential of the medium, \( e \) the charge of the electron, \( m_e \) is the mass of the electron, \( \gamma = 1/\sqrt{1 - \frac{v^2}{c^2}} \) is the Lorentz factor and \( c \) is the speed of light [3]. Depending on the energy of the photon it does
not have to be absorbed by an atomic electron and can interact with material in the ways previously discussed. The energy \( E_{\text{crit}} \) beyond which radiative effects (Bremsstrahlung) overtake ionisation as the primary form of energy loss is approximately,

\[
E_{\text{crit}} = \frac{1600}{Z} m_e c^2.
\] (2.4)

In the ATLAS electromagnetic calorimeters (lead active material) \( E_{\text{crit}} \approx 10 \text{ MeV} \) while typical electron/positron energies from a hard scattering event range from a few GeV to several hundred GeV. This means that electron energy loss at ATLAS is dominated by bremsstrahlung.

The average distance it takes a high energy electron to lose all but \( e^{-1} \) of its energy to bremsstrahlung is called the radiation length\(^3\). The radiation length gives a density independent measure of how well a material can contain high energy electrons. The radiation length of a material is approximately given by

\[
X_0 = \frac{716.4 A}{Z(Z+1) \ln\left(\frac{297}{\sqrt{Z}}\right)} \text{ [g cm}^{-2}\text{]}\] (2.5)

where \( A \) is the atomic weight of the material. By dividing \( X_0 \) by the density of the material in g cm\(^3\) a length is obtained. It can be shown [1] that the radiation length is equal to \( 7/9 \lambda \), the mean free path of photons to convert to \( e^+ + e^- \) pairs. The mean free path is defined as the distance at which there is a \( 1 - e^{-1} \) (\( \approx 63\% \)) probability that photons have converted into a \( e^+ + e^- \) pair. The probability that a photon converts to an electron-positron pair in a distance, \( x \), is given by,

\[
P(x) = 1 - \exp\left(-\frac{x}{\lambda}\right),
\] (2.6)

meaning in one radiation length there is about a 55\% chance for this to happen.

Below a few MeV, Bhabha scattering \( e^+ + e^- \rightarrow e^+ + e^- \), Möller scattering \( e^- + e^- \rightarrow e^- + e^- \) and positron annihilation become non-negligible sources of energy loss for electrons. At ATLAS the fractional energy loss due to these processes is small, only contributing near the end of a particle shower. Figure 2.3 shows the energy loss per radiation length as a function of energy.

### 2.1.3 Electromagnetic Showers

In the previous sections it was seen that the dominant form of energy loss at ATLAS for electrons is bremsstrahlung while for photons it is pair production. A simple model of an electromagnetic shower can be made to get an intuitive idea of how photons and electrons lose energy at ATLAS. The following assumptions about electrons, positrons and photons can be made:

\(^3\) \( e \) in this context is Euler’s number.
Figure 2.3: The fractional energy loss of electrons as a function of energy in lead. At low energy ionization is the dominant form of energy loss with some contribution from Bhabha scattering, Møller scattering and positron annihilation. Beyond the critical energy bremsstrahlung becomes the dominant form of energy loss [20].

1. In one radiation length, $X_0$, an electron or positron with $E \geq E_{\text{crit}}$ gives up half of its energy to a bremsstrahlung photon with no losses due to ionisation. Below $E_{\text{crit}}$ the electron or positron stops losing energy through bremsstrahlung and loses the rest of its energy through ionisation.

2. In one radiation length a photon with $E \geq E_{\text{crit}}$ undergoes pair production with each particle receiving half of the photon energy and there are no losses due to collisions with atoms. Below $E_{\text{crit}}$ photons no longer produce electron-positron pairs and lose energy entirely through collisions with atoms.

These assumptions are not true near $E_{\text{crit}}$ but at energies much higher than $E_{\text{crit}}$ and much lower than $E_{\text{crit}}$ they are approximately true. Figure 2.4 shows an electromagnetic shower in this model. Under these assumptions the approximate number of particles, $N(t)$ after $t$ radiation lengths is given by,

$$N(t) = \exp(t \ln(2)).$$  \hspace{1cm} (2.7)
Figure 2.4: A simple electromagnetic shower model where photons with energy above $E_{\text{crit}}$ lose energy entirely through pair creation and electrons/positrons with energy above $E_{\text{crit}}$ lose energy entirely through bremsstrahlung [20].

The average energy, $E(t)$, of a particle after $t$ radiation lengths relative to its incident energy, $E_0$ is given by,

$$E(t) = \frac{E_0}{2^t},$$  \hspace{1cm} (2.8)

and the radiation length, $t_{\text{max}}$ where the maximum number of particles occurs is given by

$$t_{\text{max}} = \frac{\ln\left(\frac{E_0}{E_{\text{crit}}}\right)}{\ln(2)},$$ \hspace{1cm} (2.9)

The maximum number of particles in this simple model is given by

$$N_{\text{max}} = \frac{E_0}{E_{\text{crit}}}$$  \hspace{1cm} (2.10)

and occurs when the average energy of the particles drops below $E_{\text{crit}}$ [3]. For more accurate models Monte Carlo simulations are needed and are discussed in [3].
The radial shower development is parametrised by the Molière radius,

\[ \rho_M = m_e c^2 \sqrt{\frac{4 \pi}{\alpha}} \frac{X_0}{E_{\text{crit}}} \]  \hspace{1cm} (2.11)

A cylinder with a radius of \( \rho_M \) will contain approximately 90% of the shower energy [1].

The lateral spread of an electromagnetic shower is caused by electrons and positrons scattering away from the shower axis, photons produced in the photoelectric effect and by bremsstrahlung photons emitted at a wide angle with respect to the shower axis [1].

2.1.4 Response to Electromagnetic Showers

The response of a particle is the energy the calorimeter measures divided by the true energy [1]. A calorimeter is linear if the energy measured is proportional to the energy of the particle. In linear calorimeters the response is a constant as a function of energy. In homogeneous calorimeters the entire detector volume contributes to the generated signal. These devices are linear for electromagnetic showers up to high energy. At high energy the response drops due to a number of effects such as, saturation of the photomultiplier tubes, saturation due to shower density, shower leakage or the recombination of ions with electrons inside the detector. Further calibration is required to make homogeneous calorimeters linear for electromagnetic showers at high energy.

An initial response measurement is made for photons and electrons using test beam data. The Z and J/Ψ mass peaks are used in situ to recalibrate the response and compare to Monte Carlo simulations. The energy scale at ATLAS is accurate to better than 1% for electrons and 0.3% for photons. The resolution of the response around the mean is given by the standard deviation, \( \sigma \), of the response distribution. Figure 2.5 shows that the resolution in the barrel region of the ATLAS detector is below 4% for both photons and electrons [6]. The accuracy of these measurements and the linearity provided by calibrations allow for the determination of the jet energy response to be discussed later.

![Figure 2.5](image)

Figure 2.5: The resolution of electrons (left) and photons (right) at \( \sqrt{s} = 8 \) TeV in ATLAS detector [6].
2.2 Hadronic Calorimetry

Hadrons are composite particles comprised of quarks and gluons. The strong force holds the quarks and gluons together to create three quark (baryons) and quark-antiquark (mesons) hadrons. Hadrons shower very differently in a calorimeter than purely electromagnetic particles because they experience the nuclear force. This force holds atomic nuclei together and is a residual of the fundamental strong force. In calorimeters hadrons experience a series of inelastic collisions with the nuclei of the calorimeter material resulting in a variety of interactions and decays which exhibit large event-by-event fluctuations.

2.2.1 Electromagnetic Shower Component

Hadronic particles such as the $\pi^0$ and the $\eta$ decay almost exclusively through electromagnetic interactions to two photons. These photons undergo the same showering and absorption processes that were discussed in the previous section. The energy of $\pi^0$s and $\eta$s is therefore well measured in the ATLAS calorimeter. In addition to the nuclear interactions discussed above, charged hadrons such as $\pi^\pm$ can rarely undergo bremsstrahlung and ionise atoms. Charged pions undergo bremsstrahlung rarely because they have a large mass and is the main reason that charged hadrons shower differently than electrons. These electromagnetic processes create dense electromagnetic showers within the larger hadronic shower.

2.2.2 Nuclear Interactions

Hadrons in a calorimeter strike the atomic nuclei of the calorimeter material and undergo what are known as nuclear spallation reactions. A spallation reaction occurs in a fast intranuclear cascade followed by a slower evaporation stage. In the fast stage, an incoming hadron collides with an individual nucleon within an atomic nucleus. The struck nucleon collides with other nucleons to form a cascade of nucleons which, at sufficiently high energy, can produce additional unstable hadrons. At ATLAS the energy of hadrons is usually high enough that a spallation reaction will create several hadrons. Some of these particles escape the nuclear boundary while others are trapped and excite the nucleus by distributing their energy to other nucleons.

The evaporation stage of the reaction de-excites the nucleus by releasing pions, nucleons, $\alpha$ particles and sometimes heavier nucleon aggregates, until the excitation energy is less than the binding energy of a nucleon. The energy of nucleons released in the evaporation stage is much less than the energy of particles released in the cascade. The last few MeV are released in the form of photons.

It is important to note that the cascade energy is not measured by the calorimeter. All of the energy used to overcome the nuclear binding energy is lost and this causes hadrons to have a lower response than electrons or photons. This energy is referred to as invisible.
energy. The measured energy of hadrons comes from the ionisation caused by pions and protons and evaporation neutrons. The average fractional energy deposit for a $E_0 = 1.3$ GeV pion is given in table 2.1. $E_0 = 1.3$ GeV is chosen because this is the average energy required to produce one pion in lead [7]. This value is important for later calculations but it should be noted that incident hadrons considered at ATLAS can have energies between 10-1000 GeV. At these energies roughly 90% of the particles produced in nuclear interactions are pions and the fraction of energy absorbed from hadrons through the processes of ionisation and neutron evaporation change significantly.

<table>
<thead>
<tr>
<th>Energy Component</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ionization by pions</td>
<td>19%</td>
</tr>
<tr>
<td>Ionization by protons</td>
<td>37%</td>
</tr>
<tr>
<td>Evaporation neutrons</td>
<td>10%</td>
</tr>
<tr>
<td>Nuclear binding energy loss</td>
<td>32%</td>
</tr>
<tr>
<td>Target Recoil</td>
<td>2%</td>
</tr>
<tr>
<td>Total invisible energy</td>
<td>34%</td>
</tr>
</tbody>
</table>

Table 2.1: Average fractional energy deposition of the non-electromagnetic component of a 1.3 GeV pion along with the invisible energy [1].

### 2.2.3 Hadronic Showers

The shower profile of hadrons is governed by the nuclear interaction length, $\lambda_{int}$, as well as the angular distribution of spallation reactions. The probability that a hadron traverses a distance $z$ in the calorimeter without undergoing a nuclear interaction is given by,

$$P = \exp \left( \frac{-z}{\lambda_{int}} \right),$$

so $\lambda_{int}$ is analogous to the mean free path discussed earlier for the photon. For protons in lead $\lambda_{int}$ is 170 mm compared to a radiation length of 5.6 mm for electrons. This causes hadronic showers to have a much greater depth than electromagnetic showers. Pions tend to have $\lambda_{int}$ values slightly larger than protons in most materials so hadronic showers containing a very large fraction of pions, such as those at ATLAS, are even larger. Due to the large Coulomb barrier of lead, neutrons are more abundant in hadronic showers at ATLAS than protons. The lateral profile of hadronic showers consists of a very dense electromagnetic core caused by $\pi_0$’s, that drops off at a radius of about 0.2 $\lambda_{int}$, followed by a less dense halo of hadrons. About 90% of the energy of a hadron shower is absorbed within 1.5 $\lambda_{int}$.

### 2.2.4 Response to Hadronic Showers

Hadronic showers are highly variable due to large event-by-event fluctuations in nuclear interactions. A single hadron interacting with a nucleus may turn into a few or many new
hadrons while the struck nucleus can emit varying numbers of pions, protons and neutrons, allowing for a broad range of shower properties. A typical hadron at ATLAS will produce a number of pions and since the nuclear force does not depend upon electric charge, there will be roughly a 1/3 chance of each pion being a $\pi^+$, $\pi^-$ or a $\pi^0$. The $\pi^0$'s quickly decay to photons which shower electromagnetically. Particles other than pions can be created so the probability, $f_{\pi^0}$, that a pion is created is slightly less than 1/3. The fraction of the energy in a hadronic shower carried by $\pi^0$'s after $n$ interaction iterations, $f_{em}$, is then given by,

$$f_{em} = 1 - (1 - f_{\pi^0})^n. \quad (2.13)$$

Pions can continue being produced roughly until the particles reach an energy of $E_0 = 1.3$ GeV. At each iteration, energy is converted into mass and lost to binding energy, reducing the energy available to create more particles. The number of iterations is related to the average number of particles produced in each interaction $\langle m \rangle$, also known as the multiplicity, the average energy to produce a pion $E_0$ and the incident energy $E$ by,

$$n = \frac{E}{\langle m \rangle E_0}. \quad (2.14)$$

This means that higher energy hadrons produce showers with more interactions and a larger $f_{em}$. A larger $f_{em}$ means that high-energy hadrons produce showers with a higher response.

The previous derivation ignored several important factors:

- Protons, neutrons and other hadrons are created in nuclear interactions
- The multiplicity $\langle m \rangle$, increases logarithmically with energy
- Energy loss by ionization was ignored
- Baryon number conservation was ignored

When these factors are taken into account it can be shown that,

$$f_{em}(E) = \left( \alpha_0 + \alpha_1 \ln \left( \frac{E}{E_{crit}} \right) \right) \ln \left( \frac{E}{E_{crit}} \right) \quad [21]. \quad (2.15)$$

The response of a single hadron entering the calorimeter is then given by,

$$r(E) = f_{em}(E)e + (1 - f_{em}(E))h \quad (2.16)$$

where $e$, $h$ are the response to the electromagnetic and non-electromagnetic shower components respectively [1]. The parameters $e$, $h$ are assumed to be constant as a function of energy and depend upon the composition of the calorimeter [1].

The ratio $e/h$ is used to characterise calorimeters as compensating ($e/h = 1$), under-compensating ($e/h > 1$) and overcompensating ($e/h < 1$). A compensating calorimeter is
desirable because a linear electromagnetic response implies a linear response to hadrons. There are two ways to achieve compensation; lower the response of the electromagnetic shower component or increase the response of the hadronic component.

To increase the response of the non-electromagnetic component depleted uranium can be used as an absorber material. In uranium some of the nuclear interactions create fission, releasing energy and extra neutrons. The released energy and particles compensate for the energy lost due to nuclear binding energy. Calorimeters which use fission to increase the response of the non-electromagnetic component are also referred to as compensating. ATLAS makes use of a non-compensating calorimeter with an $e/h$ value of 1.36.
Chapter 3

Experimental Setup

An overview of the Large Hadron Collider (LHC) is given in Figure 1. Figure 2 shows a diagram of the ATLAS (A Toroidal LHC Apparatus) detector and a description of the sub-detectors is given below.

3.1 The Large Hadron Collider

The LHC is a proton-proton collider located in Geneva, Switzerland. It was built with the primary objective of discovering the Higgs boson and in 2012 both the ATLAS and CMS collaborations announced the discovery of a Higgs-like boson [8]. Studies have verified that this particle is consistent with the Higgs boson predicted by the Standard Model. The LHC is a 27km circumference synchrotron accelerator. Proton bunches are accelerated using an alternating electric field while dipole magnetic fields force the protons to traverse a circular path. The proton bunches are focuced using quadrupole and higher order magnets. It is inefficient for a synchrotron the size of the LHC to accelerate particles from rest so a series of injectors, shown in Fig 3.1, is required to supply the LHC with high energy (450 GeV) protons shown. The LHC runs two beam lines in parallel but opposite directions. Collisions occur at four points corresponding to the four detectors located on the beam line: ATLAS, CMS, ALICE and LHCb.

In 2012, proton bunches were collided at a centre-of-mass energy of 8 TeV, the highest centre-of-mass energy that had ever been attained in a laboratory at the time. Bunches consisted of approximately $10^{11}$ protons with collisions occurring every 50 ns. This yielded a total integrated luminosity for both the ATLAS and CMS experiments of about 23.2 $fb^{-1}$ in 2012.
3.2 The ATLAS Detector

The ATLAS detector Fig 3.2 is a multipurpose detector which measures the properties of particles coming from proton-proton collisions in the LHC. ATLAS is the largest detector at the LHC with a length of 44 m, a diameter of 25 m and a mass of 6350 tonnes [7].

The detector must determine the charge, momentum vector and energy of particles coming from a hard scattering event. To achieve this the ATLAS detector has three layers: the inner detector, the calorimeter, and the muon spectrometer. These sub-detectors are further divided into a central (barrel) region and an end cap region. Fast electronics and software triggers are required to select events which are interesting for physics analysis. For a detailed description of the ATLAS detector see Ref [9] or [10].

3.2.1 The ATLAS Coordinate System

ATLAS analyses use the center of the detector as the origin with the z-axis parallel to the beam line, the y-axis pointing up and the x-axis pointing towards the center of the circle traced out by the LHC. Spherical coordinates are also used where,
Figure 3.2: An overview of the ATLAS detector showing the various subdetectors [9].
\[ r = \sqrt{x^2 + y^2 + z^2}, \quad (3.1) \]

\[ \phi = \arctan \left( \frac{y}{x} \right), \quad (3.2) \]

and

\[ \theta = \arctan \left( \frac{\sqrt{x^2 + y^2}}{z} \right). \quad (3.3) \]

As most particles are heavily Lorentz boosted in the z-direction, the scattering angle, \( \theta \), is replaced by pseudorapidity, \( \eta \) where,

\[ \eta = -\ln \left( \tan \left( \frac{\theta}{2} \right) \right), \quad (3.4) \]

as differences in pseudorapidity are Lorentz invariant. This allows for an angular distance measurement that is defined consistently between events,

\[ \Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}. \quad (3.5) \]

In 2012 protons in the LHC had a total momentum in the z-direction of 4 TeV but this momentum is unevenly distributed over the proton constituents. Collisions in ATLAS are energetic enough that the constituents of the protons can be resolved and a collision between two protons must be treated as a collision between a quasi-free constituent particles from each proton. The momentum in the z-direction of each particle participating in the hard scatter varies from collision to collision and therefore the momentum in the z-direction after collision is essentially never 0. Most analyses will consider the transverse components of momentum or energy for this reason.

### 3.2.2 ATLAS Inner Detector

The Inner Detector (ID) Fig 3.3 traces the path of charged particles to measure their momentum, to determine the location of primary and secondary vertices and to measure the scattering angle of the particles. The ID is immersed in a 2 T magnetic field which causes the trajectories of charged particles to bend. Based on the amount and direction of the bend, the momentum and electric charge of particles can be determined. The goal of the ID is to provide a non-destructive measurement of a particle as it travels from the interaction point to the calorimeter. Particles interact with the material of the ID at discrete points called hits which are used to reconstruct their trajectory within \( |\eta| < 2.5 \).

Fine granularity is required to resolve vertices near the interaction point. The first layer of the ID is the Silicon Pixel Detector and consists of three layers of finely segmented silicon pixels for a total of about 80.4 million read-out channels. Charged particles passing through a silicon pixel create electron-hole pairs. The electrons are collected and recorded as hits.
Surrounding the Silicon Pixel Detector is the Semiconductor Tracker (SCT) consisting of four double layered silicon strip detectors in the barrel and nine layers in the end caps. The barrel strips are long and narrow allowing the semiconductor tracker to cover a much larger area but with less granularity than the pixel detector. The layers within a strip are oriented at slightly different angles, allowing for a measurement of the position in the z direction.

The final layer of the ID is the Transition Radiation Tracker (TRT) which uses 4 mm diameter cylindrical straws (drift tubes), filled with a xenon gas mixture, to detect charged particles and photons. Charged particles ionize the gas as they pass through it creating electrons and ions. The electrons drift toward an anode wire in the center of the tube while the ions drift toward the cathode walls of the tube. Taking into account the amount of time it takes for the electrons to drift to the anode/cathode (the drift time) in the tracking algorithm allows for significantly improved spatial resolution within the TRT. This configuration does not allow for a position measurement in the z direction. The TRT also helps in particle identification by detecting transition radiation. Transition radiation is created when a relativistic charged particle passes between media of differing indices of refraction. A charged particle temporarily polarizes nearby material and as the particle passes into a material of differing refraction index the change in polarization produces electromagnetic transition radiation. The more relativistic a particle is the more transition radiation is produced and so electrons produce much more transition radiation than heavier particles. The geometry of the TRT, shown in Fig 3.4, guarantees that particles pass through at minimum 35 straws allowing for continuous tracking of particles at a larger radius than the pixel or SCT detectors.

A summary of the angular coverage and the number of channels of each subdetector in the ID is given in table 3.1.

| System       | Position         | Channels ($10^6$) | $|\eta|$ |
|--------------|------------------|-------------------|---------|
| Pixel        | Removable Barrel | 16                | $\leq 2.5$ |
|              | 2 Barrel Layers  | 81                | $\leq 1.7$ |
|              | 10 End Cap Disks | 43                | 1.7-2.5 |
| Silicon Strips| 4 Barrel Layers | 3.2               | $\leq 1.4$ |
|              | 9 End Cap Disks  | 3.0               | 1.4-2.5 |
| TRT          | Barrel Region    | 0.1               | $\leq 0.7$ |
|              | Radial Endcap    | 0.32              | 0.7-2.5 |

Table 3.1: The angular coverage and number of channels in the ATLAS ID

### 3.2.3 ATLAS Calorimeters

The ATLAS calorimeter system shown in Fig 3.5 is a non-compensating sampling calorimeter composed of sub-calorimeters which give full coverage around the beam line in the
Figure 3.3: The ATLAS Inner Detector [9]
Figure 3.4: A charged particle traverses the ATLAS Inner Detector [9].
range $0 \leq \eta \leq 4.9$. Particles entering a calorimeter interact with material to create showers of particles whose energy is absorbed and measured. Sampling calorimeters layer active material, to measure the energy of the shower particles, with dense materials to initiate particle showers. Many layers are required to fully absorb the energy from a particle at the LHC. For more information on calorimeters see the Calorimetry section. ATLAS has 5 sub-calorimeters: The Electromagnetic Barrel Calorimeter (EMB), Electromagnetic End Cap Calorimeter (EMEC), Hadronic End Cap Calorimeter (HEC), Forward Calorimeter (FCAL) and the Tile Calorimeter (TILE). These detectors are much more coarsely segmented than the ID. A Presampler for the barrel and end cap regions is located in front of these calorimeters. The presampler aims to correct for energy lost by electrons and photons before they reach the calorimeter.

The EMB is located in the barrel cryostat while the EMEC, the HEC and the FCAL are all located in the end cap cryostats. This geometry allows the ATLAS calorimeter to be thick enough to entirely absorb the energy of most particles.

The Presampler (PS) consists of an active liquid argon layer that gives a measurement of the energy of a particle and also tries to correct for energy absorbed before reaching the PS by detecting that the particle underwent an interaction in the ID. This could be due to particles interacting with dead material (e.g. support structures) or through interaction with the inner detector itself. The PS also helps to differentiate between photons and $\pi^0$’s. $\pi^0$’s decay to two photons, so two photons very close together in the highly-segmented PS likely came from a $\pi^0$. The barrel PS provides coverage from $|\eta| \leq 1.52$ while the end cap PS provides coverage from $1.5 \leq |\eta| \leq 1.8$.

The EMB and EMEC calorimeters use Liquid Argon (LAr) as the active material and lead as the passive material and are designed to accurately measure the energy of electrons, positrons and photons. Energy from charged hadrons is absorbed in these calorimeters but hadrons tend to escape the electromagnetic calorimeters. The electromagnetic calorimeters have an accordion geometry shown in Fig 3.6 which is symmetric and without cracks. The EMB provides coverage for $|\eta| \leq 1.45$ and the EMEC provides coverage for $1.375 \leq |\eta| \leq 3.2$. The EMB calorimeter consist of 3 layers of differing thickness labelled EMB1, EMB2 and EMB3. A radiation length, $\chi_0$, is the distance an electron or positron takes to lose $1 - e^{-1}$ of its energy through electromagnetic interactions. The EMB has a thickness of 22 $\chi_0$ while the EMEC has a thickness of 25 $\chi_0$ allowing for the complete absorption of most electrons, positrons and photons.

The Hadronic End Cap (HEC) sampling calorimeters consist of two independent wheels, each with two layers, behind the EMEC calorimeters. Each wheel consists of 32 identical wedge shaped modules designed to absorb the energy of hadrons. The HEC uses LAr as the active material and copper as the passive material. The HEC extends from $1.5 \leq |\eta| \leq 3.2$, slightly overlapping both the Forward and Tile Calorimeters. The HEC calorimeters have a combined thickness of about 10 interaction lengths, $\lambda$, shown in Fig 3.7.
Figure 3.5: The ATLAS Calorimeter [9]
The Forward Calorimeter (FCAL) provides coverage from $3.2 \leq |\eta| \leq 4.5$ and consists of 3 layers. This high-radiation region of the detector must use different materials to absorb high-energy jets. The first layer is LAr and copper and is optimised for particles that interact electromagnetically while the next two layers are LAr and tungsten, and mostly measure the energy of hadrons. The high density tungsten reduces punch through into the muon spectrometer and mitigates leakage into the HEC. In total the FCAL has a thickness of about $10 \lambda$.

The hadronic Tile calorimeter (TILE) uses steel as the passive material and plastic scintillator as the active material. It covers the region $|\eta| \leq 1.7$. The TILE is designed to absorb all of the energy from jets after they have passed through the EMB ($\approx 1.2 \lambda$ in thickness). The plastic tiles are staggered with depth and are attached to photomultipliers as shown in figure 3.8. An electronic pulse shaper interprets the current pulse from the photomultiplier as signal in 50 ns. The TILE calorimeter is broken into a barrel, covering
$|\eta| \leq 1.0$, and extended barrel region covering $0.8 \leq |\eta| \leq 1.7$. Each have three layers but the barrel layers, TILEB1, TILEB2 and TILEB3 are of special interest to the work described in this thesis as only jets with $|\eta| \leq 0.8$ are considered.

### 3.2.4 Muon Spectrometer

High-energy muons primarily lose energy through ionization and bremsstrahlung. At the energies muons are produced at ATLAS the energy loss due to ionization is approximately minimized (known as a minimum ionizing particle). Meanwhile energy loss due to bremsstrahlung is proportional to $\frac{1}{m^2}$ where $m$ is the mass of the particle. As muons are fairly massive, the energy loss due to bremsstrahlung is small. Muons at ATLAS typically pass through the calorimeters and require a dedicated detector to measure their energy. The muon spectrometer Fig 3.9 makes use of large toroidal magnets to bend muons in the $r$-$z$ plane. The magnetic field is divided into the barrel $|\eta| \leq 1$, transition $1 \leq |\eta| \leq 1.4$ and end cap $1.4 \leq |\eta| \leq 2.7$ regions with differing magnetic field strengths. Four subsystems: The Monitored Drift Chambers (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) are used to identify and track muons as they pass through the magnetic field. The MDT is a drift tube chamber while the CSC is a multiwire proportional chamber. In both systems charged particles ionize an Argon-CO2 gas mixture releasing electrons. The electrons drift to an anode wire generating a
current signaling that a muon has passed. Positive ions are also used to generate the signal for a muon. The RPC and TGC are mostly for triggering on muons but also give some measurement of energy. Instead of wires these two systems read out charged plates.
Figure 3.9: The Muon spectrometer subdetectors [9]
Chapter 4

Physics Object Reconstruction

Event reconstruction is the process through which detector signals are combined and re-assembled to describe physical objects. Physical objects include particles, such as electrons, muons or photons, composite objects like jets, as well as Missing Transverse Energy (MET). Further analyses can be used to identify, \( \tau \)'s, top quarks and b quarks.

4.1 Calorimeter Clusters

Calorimeter clusters are built from single readout channels in the calorimeter called cells. A clustering algorithm looks for energy deposits in cells and combines them into clusters. There are two main clustering algorithms used at ATLAS; sliding window clusters and topological clusters (Topoclusters) [11].

4.1.1 Sliding Window Clusters

The sliding window algorithm is used to identify electrons and photons. Cone shaped three-dimensional *towers* in the detector are constructed using a window of area \( \Delta \eta \times \Delta \phi \). Preclusters are formed by moving the window around the calorimeter to find local transverse energy maxima above a 3 GeV threshold [11]. For electrons and photons the precluster window size is 5 cells x 5 cells. This size and threshold have been optimised to obtain the best efficiency for finding preclusters while limiting the rate of fake preclusters. The position of the precluster is calculated using the energy weighted barycenter of cells within a 3x3 sub-window of the precluster. Using a smaller window makes the position calculation less susceptible to noise [11].

Preclusters are used as seed clusters in searching for particles. Windows are placed around the seed clusters to capture the energy of particles. The window size depends upon the particle hypothesis and the location in the detector and is optimized to contain most of the particle energy while limiting noise (see Table 4.1). The window size for photons depends on whether the photon converts into an electron-positron pair within the tracker.
(converted photon) or does not (unconverted photon). The window size is slightly larger for converted photons in the barrel because the magnetic field permeating the tracker will force the electron and positron to bend in opposite directions.

<table>
<thead>
<tr>
<th>Particle Type</th>
<th>Barrel</th>
<th>End Cap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron</td>
<td>3 × 7</td>
<td>5 × 5</td>
</tr>
<tr>
<td>Converted Photon</td>
<td>3 × 7</td>
<td>5 × 5</td>
</tr>
<tr>
<td>Unconverted Photon</td>
<td>3 × 5</td>
<td>5 × 5</td>
</tr>
</tbody>
</table>

Table 4.1: Window dimensions (η × φ) for electrons, converted photons and unconverted photons [11].

### 4.1.2 Topological Clusters

The topological clustering algorithm groups cells together into topoclusters based on how significant their energy is compared to noise levels. Three-dimensional topoclusters are used as seeds for jet reconstruction. Topoclusters have a variable size and shape which allows them to capture most of the energy from single particles.

Cells with an energy-to-noise ratio above some threshold $t_{\text{seed}}$ are used as seed cells for this algorithm. The noise can come from a combination of electronic noise, other activity in the same beam crossing (in-time pile-up) or activity from an adjacent beam crossing (out-of-time pile-up). Cells surrounding the seed cells with a signal-to-noise ratio above $t_{\text{neighbour}}$ are combined with the seed cell and new neighbours are checked. Finally, any cells above a threshold of $t_{\text{cell}}$ surrounding the cluster are added to the cluster. The algorithm is usually referred to as $t_{\text{seed}}/t_{\text{neighbour}}/t_{\text{cell}}$. ATLAS uses a 4/2/0 topoclustering algorithm to build jets. The 0 in this case means that there is no requirement on the outer cells, which are added to the cluster. The high $t_{\text{seed}}$ threshold reduces the rate of fake topoclusters while the low threshold $t_{\text{cell}}$ ensures that all of the energy from a particle is captured.

### 4.2 Jets

In a hard scattering event, colliding protons interact through individual collisions between quarks and gluons. The interacting quarks and gluons gain enough energy and momentum to break the proton apart, initiating parton showers. In a parton shower, gluons with enough energy can split into a quark-antiquark pair. Because gluons carry color themselves, they can also radiate gluons. Quarks from gluon splitting or the initial hard scattering can radiate more gluons. Eventually the partons lose enough energy that stable, colour neutral hadrons can be formed. This last step is a non-perturbative process that cannot be calculated analytically. A model is therefore used. The entire process is known as fragmentation and results in a collimated spray of hadrons travelling roughly in the direction of the original
parton. This shower of hadrons travels through the detector until it reaches the calorimeter at which point calorimeter showers are created as discussed in section 2.2.

The particles originating from a single parton immediately following the fragmentation process are referred to as the particle jet. The calorimeter energy deposits which these particles produce form the calorimeter jet. Either of these objects may be called a jet and the meaning must be inferred from the context.

The entire process from hard scattering to calorimeter jet is very complicated with multiple steps. Many algorithms have been developed to reconstruct jets from particles or their calorimeter energy deposits. These jet algorithms can be applied at various stages of jet development and may be used on a variety of inputs. Jets can be constructed at three stages; at parton level from partons before fragmentation, at particle level from particles immediately after fragmentation, and at calorimeter level from calorimeter clusters.

Jet algorithms need to satisfy a number of criteria in order to be used in physics analysis. The algorithm should be easy to implement, make efficient use of computer resources, reconstruct the same jets regardless of Lorentz boost and find the same jets at parton, particle and calorimeter level. The algorithm should be insensitive to low pT (soft) radiation as the number of soft gluons emitted diverges in a fixed order QCD calculation. This is a property known as Infrared Safety. Infrared safety means that adding a soft particle should not cause two jets to be merged in the reconstruction algorithm. Jet algorithms should be collinear safe, meaning they should produce the same jets regardless of whether a parton splits into two partons travelling in the same direction. For a more detailed discussion of jet algorithms see Ref. [12].

An algorithm that satisfies these criteria and is used by ATLAS is the anti-kt algorithm [13]. In this algorithm all relevant objects are considered to be pseudo-jets. A distance parameter, $d_{i,j}$ is calculated to measure the distance between each pair of pseudo-jets $i$ and $j$. $d_{i,j}$ is defined for the anti-kt algorithm by,

$$d_{i,j} = \min \left( p_{T,i}^{-2}, p_{T,j}^{-2} \right) \frac{\Delta_{ij}^2}{R^2},$$

where $p_{T,i}$ is the transverse momentum of pseudo-jet $i$, $\Delta_{ij}^2$ is the distance in $\eta\phi$ space between pseudo-jets $i$ and $j$ and $R$ is a size parameter. A distance parameter between each pseudo jet and the beam line is given by,

$$d_{i,B} = p_{T,i}^{-2}.$$  

If the smallest $d$ is a pseudo-jet pair then the two pseudo-jets are merged. If the smallest $d$ is a single pseudo-jet then it is moved to the list of jets. This process is repeated until no pseudo-jets remain. Jets for this thesis are created from topoclusters using the anti-kt algorithm with a size parameter of 0.6.
4.3 Muons

The first step in muon reconstruction at ATLAS is to fit straight lines, called track or muon segments, to hits in the muon spectrometers. Track segments from the outermost muon spectrometer layer are then matched to segments in the middle and inner layers until a full muon track in the muon spectrometer is reconstructed. After the reconstruction the muon tracks are extrapolated back to the interaction point. This so-called standalone reconstruction is done using the Muonboy algorithm.

In the region $|\eta| \leq 2.5$, further analysis can be done to match muon tracks to tracks in the Inner Detector (ID). Staco combined muons are reconstructed using a statistical combination of muons reconstructed from ID hits and muons reconstructed from muon spectrometer hits. An alternative algorithm is MUID which uses a global fit to muon hits in both the ID and the muon spectrometer simultaneously [10].

4.4 Missing Tranverse Energy

In physics analyses it is often useful to measure the imbalance of momentum produced in a collision. A large imbalance can indicate a high energy particle, such as a neutrino, escaping detection. On the other hand a small imbalance can be produced by a mismeasurement. As discussed previously, due to the proton’s composite nature the total momentum in the z-direction associated with the hard scatter is not 0. Ignoring intrinsic transverse momentum in the proton and beam misalignment, the momentum in the transverse plane is guaranteed to be 0. For this reason missing transverse energy, $E_{\text{miss}}$ is usually considered.

$E_{\text{miss}}$ can be constructed from a variety of objects ranging from cells or clusters, to tracks, to fully reconstructed jets. In this thesis $E_{\text{miss}}$ has been reconstructed using topoclusters. $E_{\text{miss}}$ values in the x,y-directions are calculated by summing over the missing energy in the specified direction of the clusters,

$$E_{\text{miss}}^{T_{x,y}} = \sum E_{\text{cluster}}^{T_{x,y}}. \quad (4.3)$$

The full missing energy is obtained by summing in quadrature $E_{\text{miss}}^{T_{x,y}}$.

Muons complicate the situation since they only deposit a small fraction of their energy in the topoclusters. To correct for this two approaches are used. In the first approach the muon momentum from the fully reconstructed muon is calculated and then an estimate to the amount of energy the muon deposited in the calorimeter is made. This energy is subtracted from $E_{\text{miss}}^{T}$ to avoid double counting it. In the second approach the muon momentum used to calculate $E_{\text{miss}}^{T}$ is the momentum measured by the muon spectrometer alone. The difference between the momentum measured by the muon spectrometer and the momentum of the fully reconstructed muon is precisely the momentum deposited in the
calorimeter. Therefore using the momentum measured by the muon spectrometer accounts for all the energy without double counting.

4.5 Triggers

In 2012 the LHC produced more than 20 million interactions per second in the ATLAS detector, each requiring about 2 megabytes to store. To reduce the amount of storage space and computation power required to process the events only a selection of events is recorded. ATLAS uses a 3 level trigger system to reduce the 20 million interactions per second down to a few hundred events of interest per second. The first level trigger is the L1 trigger. It is a hardware based trigger which makes decisions based on the number of particles above a certain momentum. The L1 trigger associates these triggered objects to a region of interest in the detector. The second level trigger is software based. It makes decisions based on reconstructed and calibrated objects within the regions of interest. Objects outside of the region of interest are not reconstructed or calibrated to save time. Calibrated objects which pass the L2 trigger are sent to the third level, event filter (EF) trigger. The EF trigger uses fully reconstructed and calibrated events to make a final series of selection requirements. The L2 and EF level triggers are implemented on large computer clusters.

To further reduce the number of events of a certain type, only a fraction (the prescale factor) of events are kept in a process called prescaling. After passing the selection criteria of a trigger, a fixed fraction of events are chosen at random to be used in analysis. This sub-sample of events has been ‘prescaled’ from the full set of events which pass a trigger and give an unbiased representation of the whole. To retrieve the total number of events the prescaled sample can be weighted by the prescale factor.
Chapter 5

Jet Energy Scale

ATLAS measures the energy of jets at calorimeter level and uses the jet energy scale (JES) calibration to extrapolate the jet energy to particle level. This allows experimental data to be compared to theory calculations that have been propagated from generator level using a model for quark fragmentation. The response of a jet is the ratio of the measured energy of the jet to its particle level energy. It is important to minimize the uncertainty of the response because it is a large component of the total uncertainty for many analyses involving jets. The JES correction and uncertainty (red in figure 5.2) are calculated using in-situ techniques on specific data sets. Any other datasets must have additional uncertainties to take into account the varying flavour composition of the different process (blue in figure 5.2) as well as the flavour dependant JES. The JES flavour uncertainties (green in figure 5.2) are due to the fact that quark-initiated jets and gluon-initiated jets are slightly different. Gluon jets tend to have a larger number of particles, each with lower energy, than in quark-initiated jets. This causes gluon initiated jets to have a lower response than quark initiated jets. In 2011 the response difference between quark and gluon EM scale anti-kt 0.4 jets was estimated using different Monte Carlo generators to be between 6-8% at low $p_T$. It decreases as $p_T$ increases as shown in figure 5.1. The relative JES uncertainty for antiKt 0.4 jets assuming a jet composition of 50 % quark and 50 % gluon is shown in figure 5.2.

The in-situ-correction mostly comes from samples with a large light-quark fraction and therefore any sample with a gluon fraction greater than 50 % (ex. multi-jet) will have a larger flavour uncertainty than in figure 5.2. The goal of this thesis is to measure the response of quark and gluon jets in data in the hope of reducing the JES flavour uncertainty. Before this is accomplished the ATLAS JES calibration framework should be discussed.

5.1 ATLAS JES Calibration Framework

ATLAS uses two types of calorimeter-based jet calibrations, Electromagnetic Scale (EM) jets and Local Cluster Weighting (LCW) jets. EM scale sets the correct cluster energy for
Figure 5.1: 2011 response difference in Monte Carlo simulations between quark and gluon jets measured at EM scale using the antikt 0.4 algorithm [14].

Figure 5.2: 2011 relative jet energy scale uncertainties as a function of $p_T$ for $\eta = 0.5$ (left) and $\eta = 2.0$ (right) for jets measured at EM scale using the antikt 0.4 algorithm. A 50 % quark and 50 % gluon composition was assumed [14].

electrons and photons with no cluster energy correction for hadrons. EM jets are built from these clusters using the anti-kt algorithm described in the previous chapter. Clusters may contain energy from both electromagnetic and hadronic showers. Local Cluster Weighting uses information about the cluster to determine if the energy within a cluster appears electromagnetic or hadronic and applies an energy correction based on this. LCW jets are
then built from the energy corrected clusters. In this thesis only EM jets will be studied but the principles described also apply to LCW jets.

Figure 5.3: Jets are built from calorimeter clusters at either EM or LCW scale followed by an energy calibration in several steps [14]

Once jets are constructed the calibration procedure begins. The first step is to correct for energy from pile-up. Pile-up includes all particles that do not come from hard scattering processes. When bunches of $\approx 10^{11}$ protons collide it is possible that multiple interactions occur. In 2012 the average number of interactions per bunch crossing was 21 [15]. Most of these interactions are glancing, low $p_T$ interactions which are not interesting for physics analyses but contribute to background noise. This type of pile-up is referred to as in-time, meaning that it comes from the same bunch crossing as a hard scattering event. Out-of-time pile-up comes from bunch crossings which occur either before or after the beam bunch crossing with the hard scattering event. Out-of-time pile-up affects the energy measurement of jets because the LAr calorimeters have a large response time of a few hundred ns.

Pile-up energy subtraction occurs in two steps, the first of which is an area-subtraction. In this step an “area” for all jets (both hard scattering jets and pile-up jets) is calculated using the FastJet algorithm. Infinitesimally low energy ghost particles are uniformly distributed over the detector and included in the jet clustering algorithm. The number of ghost particles associated with a jet is a measure of the jet area. This process is somewhat complicated but for anti-kt jets the area is roughly $\pi R^2$ as expected for a circular anti-kt jet as long as it does not overlap with another jet. The advantage of this process is that it allows for a consistent area definition for non-circular jets produced by different jet algorithms. The median $p_T$ density in an event, $\rho$, is then defined as

$$\rho = \text{median}\left\{\frac{p_T^{\text{jet}_i}}{A_i}\right\}, \quad (5.1)$$
where $p_{T,i}^\text{jet}$ is the $p_T$ of jet $i$ and $A_i^\text{jet}$ the area. As a first pass the corrected transverse momentum of a jet, $p_T^\text{corr}$, is given by

$$p_T^\text{corr} = p_T - \rho \times A \ [15].$$

(5.2)

A residual correction is then applied based on the number of primary vertices (NPV) in the bunch crossing and the average number of interactions per bunch crossing, $\langle \mu \rangle$. NPV is directly related to the in-time pile-up while $\langle \mu \rangle$ gives a measure of both in-time and out-of-time pile-up. The final pile-up corrected energy is given as,

$$p_T^\text{corr} = p_T - \rho \times A - \alpha \times (NPV - 1) - \beta \times \langle \mu \rangle,$$

(5.3)

where $\alpha$ and $\beta$ are linear approximations to the rate of change of pile-up $p_T$ with respect to NPV and $\langle \mu \rangle$, respectively [15].

Pile-up jets are suppressed in analyses by making use of the Jet Vertex Fraction (JVF) variable. JVF uses charged tracks (Energy $\geq$ 500 MeV) from primary vertices (PV) to measure the fraction of the jet energy that comes from a specific PV (figure 5.4). A cut is placed on the JVF to reject jets that are dominated by pile-up. JVF is calculated for each jet $i$ with respect to each PV$_j$ as

$$\text{JVF}(\text{jet}_i, \text{PV}_j) = \frac{\sum_k p_T(\text{track}_{k,jet}^i, \text{PV}_j)}{\sum_n \sum_l p_T(\text{track}_{l,jet}^i, \text{PV}_n)},$$

(5.4)

where $k$ runs over all tracks matched to jet $i$, $n$ runs over all PVs in the event and runs $l$ over all tracks originating from PV$_n$. In 2012 the JVF cut applied to leading and subleading jets was $\geq 0.25$.

The next step is an $\eta$ correction which does not change the energy of the jet but adjusts the direction of the jet to point from the primary vertex to the jet centroid. This is because the PV is likely not at the origin of the ATLAS detector. The initial Monte Carlo JES correction, which is validated by test beam, is applied at this point.

In the final step an in-situ correction is applied. Several techniques exist to derive this calibration but in general momentum balance in the transverse plane is exploited. Events where a jet recoils against a well measured object, such as a photon or Z boson, are used to correct the energy of the jet in the central region. Once the energy of jets in the central region is calibrated, dijet (multijet) events can be used to extend the calibration to higher $\eta$ (energy).

### 5.2 Jet Response

The largest component of the JES correction is the absolute jet response. The jet response measures how the calorimeter responds to both electromagnetically and hadronically in-
teracting objects within a jet. Recall that for a single hadron entering the calorimeter the response was given by,

\[ r(E) = f_{em}(E)e + (1 - f_{em}(E))h, \]  

(5.5)

where \( e \) and \( h \) are the respective responses to the electromagnetic and hadronic shower components. The response of a jet is then given by,

\[ j(E) = w_h \cdot r(w_h \cdot E) + w_e \cdot e(w_e \cdot E), \]  

(5.6)

where \( E \) is the energy of the jet and \( w_h \) and \( w_e \) are respectively the fractions of the particle jet that decay hadronically and electromagnetically. Two assumptions must be made to arrive at the calorimeter jet response. First, it must be assumed that the responses to the electromagnetic and hadronic components (\( e \) and \( h \)) are energy independent. The second assumption is that the fraction of electromagnetically decaying particles at particle level is also energy independent. These two assumptions allow the jet response to be parametrized by,

\[ j(E) = b_0 + b_1 \cdot \ln \left( \frac{E}{E_{scale}} \right) + b_2 \cdot \ln^2 \left( \frac{E}{E_{scale}} \right), \]  

(5.7)

where recall \( E_{scale} = 1.3 \text{ GeV} \) [1].

5.3 In-situ Techniques

The three in-situ techniques used to measure the response of jets at ATLAS are Direct Balance (DB), Missing-\( E_T \) Projection Fraction (MPF) and Multi-Jet Balance (MJB). All three
techniques use a reference object that is roughly back-to-back with a jet in the transverse plane.

The Direct Balance technique uses the ratio of jet $p_T$ to the reference object $p_T$. The reference object in this technique is either a photon or a $Z$ boson. In order to reduce the effect of parton radiation perpendicular to the jet axis only the $p_T$ component in the direction of the jet is taken. This gives $p_T^{ref}$, as,

$$p_T^{ref} = p_T^{\gamma/Z} \times \cos (\Delta \phi (\gamma/Z, jet)), \quad (5.8)$$

where $\Delta \phi (\gamma/Z, jet)$ is the azimuthal angle between the reference object and the jet. The DB response is given as,

$$R_{DB} = \frac{p_T^{jet}}{p_T^{ref}} \quad (5.9)$$

The MPF technique makes use of the missing transverse energy (MET) in a $\gamma/Z +$ jet event to measure the response of a jet. To derive $R_{MPF}$, the response of the hadronic recoil, momentum balance at parton level is first considered. At parton level, momentum balance in the transverse plane is guaranteed for $\gamma/Z +$ jet events,

$$\vec{p}_{parton}^T + \vec{p}_{\gamma/Z}^T = 0 \quad (5.10)$$

Under the assumption that fragmentation does not affect momentum balance at particle level, it can be assumed that the particle level jet balances the reference object (the $\gamma$ or the $Z$),

$$\vec{p}_{recoil}^T + \vec{p}_{\gamma/Z}^{ref} = 0 \quad (5.11)$$

Selection criteria on event topology are used to suppress both initial and final state radiation to ensure the hadronic recoil balances the reference object. At calorimeter level momentum balance reads,

$$\vec{p}_{meas, recoil}^T + \vec{p}_{meas, ref}^T = -\vec{E}_{miss}^T \quad (5.12)$$

where $\vec{p}_{meas, recoil}^T = R_{MPF} \cdot \vec{p}_{recoil}^T$, $\vec{p}_{meas, ref}^T = e \cdot \vec{p}_{\gamma/Z}^{ref}$ and the minus sign emphasizes the fact that the missing energy is in the opposite direction to the energy excess. Substituting these into the above equation, noting that $e \approx 1$

$$\vec{p}_{\gamma/Z}^{ref} + R_{MPF} \cdot \vec{p}_{recoil}^T = -\vec{E}_{miss}^T, \quad (5.13)$$

The difference in the EM response and the jet response is a source of missing energy in the direction of the jet. The $\vec{E}_{miss}^T$ due to pile-up, the underlying event and other sources is assumed to be $\phi$-symmetric with respect to the hadronic recoil/reference object system. Therefore, over many events the $\vec{E}_{miss}^T$ due to these other sources averages to zero, leaving
only the $\vec{E}_T^{\text{miss}}$ due to the calorimeter response. By projecting along the direction of the reference object and using momentum balance at particle level this can be rewritten as,

$$p_T^{\text{ref}} - R_{MPF} p_T^{\text{ref}} = -\vec{n}_{\text{ref}} \cdot \vec{E}_T^{\text{miss}}, \quad (5.14)$$

where $\vec{n}_{\text{ref}}$ is the unit vector in the direction of the reference object. Rearranging, the MPF response, $R_{MPF}$, is given by

$$R_{MPF} = 1 + \frac{\vec{n}_{\text{ref}} \cdot \vec{E}_T^{\text{miss}}}{p_T^{\text{ref}}} . \quad (5.15)$$

The third in-situ method is the MJB. The previously discussed methods are used to calibrate jets up to 800 GeV, beyond which there are not enough $\gamma/Z + \text{jet}$ events to perform the calibration. To calibrate jets to higher energies, multi-jet events are used. The leading jet is balanced against a system of lower energy jets that have been calibrated using DB and MPF. The MJB method is compared to the $\gamma + \text{jet}$ response from DB and MPF in the energy range 400-800 GeV.

This thesis focuses on the MPF technique to measure the jet response in $\gamma + \text{jet}$ events and, after applying an additional correction, dijet events. This correction is the subject of chapter 6.

### 5.4 Quark and Gluon Jets

Quarks and gluons are governed by the strong force described by Quantum Chromodynamics (QCD). An additional quantum number, colour, is required to satisfy the Pauli exclusion principle. Quarks come in three colours: red, green and blue while anti-quarks have anti-colour: anti-red, anti-green and anti-blue. Gluons exist in one of 8 states that are combinations of colour and anti-colour. The allowed states come from the SU(3) colour-octet fundamental to QCD.

QCD exhibits asymptotic freedom, meaning that at high energy the strong force is weak and allows theoretical calculations to be performed perturbatively. High energy interactions are mediated by single gluon exchange while Monte Carlo simulations must be used to calculate the probability for interaction of quarks and gluons in non-perturbative regimes. The strength of the strong coupling for single gluon exchange between two colour charges is given by $\frac{1}{2}c_1 c_2 \alpha_s$ where $c_1$, $c_2$ are the colour coefficients associated with the vertices and $\alpha_s$ is the strong coupling coefficient. The colour factor $C_F$ is defined as,

$$C_F = \frac{1}{2} |c_1 c_2| \quad (5.16)$$

and is used to compare the strength of different processes. The allowed processes in single gluon exchange are “quark radiates a gluon”, “gluon splits into quark-antiquark pair” and
“gluon radiates a gluon”. The colour averaged-colour factors for these processes are given in table 5.1 where a higher colour factor indicates a higher probability of occurring.

<table>
<thead>
<tr>
<th>Process</th>
<th>Symbol</th>
<th>Colour Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quark Radiates Gluon</td>
<td>$C_F$</td>
<td>$\frac{4}{3}$</td>
</tr>
<tr>
<td>Gluon splitting</td>
<td>$T_F$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>Gluon Radiates Gluon</td>
<td>$C_A$</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5.1: Colour factor for single gluon exchange interactions [16]

The ratio $\frac{C_A}{C_F} = \frac{9}{4}$ implies that gluons radiate more particles than quarks. During fragmentation this means gluon jets develop more particles than quark jets as shown in 5.5.

Figure 5.5: A depiction of how the colour factor affects jet structure during fragmentation, resulting in gluon jets producing more particles than quark jets [30].

In gluon jets, the energy from the hard scattering is spread out over more particles meaning that on average each particle within the jet has lower energy than in quark jets. This has a couple of effects on jet structure.

1. Gluon jets tend to have a higher particle multiplicity than quark jets (1-3 more tracks [17])
2. Gluon jets tend to be broader than quark jets
3. Quark jets tend to penetrate deeper into the calorimeter than gluon jets
4. Quark jets have a denser energy profile than gluon jets

Gluon jets tend to have lower response than quark jets because of their large number of lower energy particles. As previously discussed, the response to a hadronic shower depends upon its energy, with lower energy showers having a worse response. Low energy particles within a shower also have a higher probability of being bent outside of the jet by magnetic fields, causing the response of a jet to be lower. As gluon jets tend to contain lower energy particles this affects gluon jets more than quark jets and lowers their response relative to quark jets [17].
Monte Carlo simulations give an estimate of the response difference between quark and gluon jets but this has yet to be measured in data. Attempts have been made to tag quark and gluon jets in data using the previously discussed differences [18] although a response measurement was not made.

Separate light-quark and gluon jet responses could be used to derive a jet response for an arbitrary data sample through the formula,

$$R_S = R^q \times f^q_S + R^g \times f^g_S,$$  \hspace{1cm} (5.17)

where $R_S$ is the response of jets in the sample, $R^q$, $R^g$ are the response of quark and gluon jets and $f^q_S$, $f^g_S$ are the fractions of light-quark and gluon jets in the sample. The assumption in the above equation is that the fraction of jets coming from c and b quarks is negligible. For non-negligible c and b fractions, estimates from Monte Carlo simulations can be used.

In order to calculate a light-quark and gluon response it is possible to use the response of jets in dijet and photon + jet events along with their respective quark and gluon jet fractions to obtain a quark and gluon jet response. The dijet response, $R_{DJ}$, may be written as,

$$R_{DJ} = R^q \times f^q_{DJ} + R^g \times f^g_{DJ} + R^c \times f^c_{DJ}, \hspace{1cm} (5.18)$$

where $f^q_{DJ}$, $f^g_{DJ}$ and $f^c_{DJ}$ are the light quark, gluon and c quark jet fractions in dijet events and $R^q$, $R^g$ and $R^c$ are their respective responses. Similarly for photon + jet events,

$$R_{\gamma J} = R^q \times f^q_{\gamma J} + R^g \times f^g_{\gamma J} + R^c \times f^c_{\gamma J}. \hspace{1cm} (5.19)$$

Bottom quarks contribute a negligible amount for the dijet and photon + jet responses. The quark and gluon fractions as well as $R^c$ are estimated using Monte Carlo simulation. Solving these equations for the quark and gluon responses gives,

$$R^q = \frac{f^q_{DJ} \left( R_{DJ} - R^c f^c_{\gamma J} \right) - f^q_{\gamma J} \left( R_{\gamma J} - R^c f^c_{DJ} \right)}{f^q_{\gamma J} f^q_{DJ} - f^q_{DJ} f^q_{\gamma J}}, \hspace{1cm} (5.20)$$

and

$$R^g = \frac{f^g_{DJ} \left( R_{DJ} - R^c f^c_{\gamma J} \right) - f^g_{\gamma J} \left( R_{\gamma J} - R^c f^c_{DJ} \right)}{f^g_{\gamma J} f^g_{DJ} - f^g_{DJ} f^g_{\gamma J}}. \hspace{1cm} (5.21)$$

The challenge is to measure the response of jets in dijet events. To accomplish this a correction will be developed to allow one of the jets to be used as a reference object in the MPF technique.
Chapter 6

Jet Properties and Response Correction

The goal is to provide a well measured reference object for the MPF technique in dijet events. In dijet events the width of the response distribution in a given $p_T$ range is broad. The Monte Carlo response, $R_{MC}$, is defined as,

$$R_{MC} = \frac{Jet_{pT}^{\text{PileCorr}}}{Jet_{pT}^{\text{Truth}}}$$

(6.1)

where $Jet_{pT}^{\text{PileCorr}}$ is the pile-up corrected transverse momentum measured by the calorimeter and $Jet_{pT}^{\text{Truth}}$ is the transverse momentum at particle level. This is shown in figure 6.1 for the transverse momentum range 105 GeV to 125 GeV.

To get an estimate of the particle level jet $p_T$ in data, for a given $p_T$ range, the $Jet_{pT}^{\text{PileCorr}}$ could be divided by the mean of the response distribution in that $p_T$ range. Correcting the jet $p_T$ in this way would ensure that jets have the true particle level $p_T$, on average, but an individual jet’s $p_T$ could be far from the true value. This is because the response distribution in dijets is very broad as shown in 6.1. The $\sigma$ of the response distribution gives an estimate of the $p_T$ error when correcting jet energy in this way.

The goal is to find a way to minimize the width of the jet response distribution. It turns out that the response of jets depends upon certain jet properties, such as, jet shape, the energy distribution of a jet in the calorimeter and the attributes of charged tracks within the jet. A Monte Carlo simulation study was undertaken to determine the relationship between jet variables and jet response.

6.1 Jet Variable Study

To analyse the relationship between jet variables and response, jets were divided into two groups; a high response group and lower response group. For the main analysis the high response group was jets with the top 10% response and the lower response group was jets
with the bottom 70% response in a given momentum range. Alternative definitions were explored using a 10% high/50% low, 20% high/70% low and 20% high/50% low for the high response and low response groups. These alternative definitions are used to determine a contribution to the systematic uncertainty on the method.

The relationship between jet properties and jet response is explored using two different Monte Carlo simulations Pythia8 and Herwig++. Both Pythia8 and Herwig++ simulate 2-2 (two incoming and two outgoing) QCD interactions at next-to-leading order. The two simulation models hadronise the partons resulting from their respective parton showers. Pythia8 uses a string fragmentation model while Herwig++ uses a cluster fragmentation model [22, 23]. In the string fragmentation model all gluons are shared between a quark-antiquark pair, represented by the long tubes in Figure 6.2. The quark-antiquark pair along with the gluons shared between them form a colour neutral state that (usually) decays into hadrons. For a thorough discussion of the string fragmentation model see Ref [24]. In a cluster model, gluons at the end of the parton shower are split into colour-connected quark-antiquark pairs. The quark from a gluon splitting is paired with a (possibly different flavour) anti-quark from another gluon splitting to form a cluster. Clusters are only made from singlet state quark-antiquark pairs \(^1\). Clusters are then randomly assigned a quark-

\(^1\)colour singlet states are \(r\bar{r}, b\bar{b}\) and \(g\bar{g}\).
antiquark pair or a diquark-diantiquark pair based on phase space availability. The clusters then decay into either two mesons or a baryon and anti-baryon [19, 26].

![Diagram](image)

Figure 6.2: The cluster (left) and string (right) fragmentation models [19]

The jet properties analysed are given in table 6.1. Figure 6.3 shows the variable distributions for select variables in the $p_T$ range 105-125 GeV for data (black), Pythia (red) and Herwig++ (blue). All variable distributions in the $p_T$ ranges studied are shown in appendix D. The differences between high response and low response jets come partly from fragmentation and partly from calorimeter showering. Parton showers which fluctuate to have a high $\pi^0$ content will have a high response because $\pi^0$'s decay almost exclusively to two photons on a very short time scale. This creates an electromagnetic component within a jet, the energy of which is measured accurately.

Hadronic showers in the ATLAS calorimeter are largely composed of $\pi^0$, $\pi^+$ and $\pi^-$. The nuclear force treats each of these particles in the same way so there is an equal probability of creating each type of pion in a nuclear interaction. High response jets are formed when calorimeter showers fluctuate to have a large fraction of $\pi^0$'s early on in the showering process, creating a large electromagnetic jet component. The energy in the electromagnetic component of the jet is fully absorbed by the detector giving the jet a high response.

Jets with a high $\pi^0$ content will be more similar to electromagnetic showers than jets with a low $\pi^0$ content. High response jets will, on average, have a higher energy density, fewer charged tracks and have a smaller physical size than low response jets.
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTrkX</td>
<td>The number of tracks with at least X = 500 MeV, 1 GeV momentum within the jet</td>
</tr>
<tr>
<td>TrackPtXFrac</td>
<td>The fraction of the jet momentum carried by charged tracks with at least X = 500 MeV, 1 GeV momentum</td>
</tr>
<tr>
<td>AVGTrackPtXFrac</td>
<td>The fraction of the jet momentum carried by charged tracks with at least X = 500 MeV, 1 GeV divided by the number of tracks with at least that energy</td>
</tr>
<tr>
<td>TrackWidth</td>
<td>The width of a jet using tracks with at least 1 GeV energy. The width gives a measure of energy density through the formula $W = \frac{\sum_i \Delta R^i \cdot p_T^i}{\sum_i p_T^i}$</td>
</tr>
<tr>
<td>Width</td>
<td>The width of a jet calculated using calorimeter clusters instead of tracks</td>
</tr>
<tr>
<td>PREBFrac</td>
<td>The fraction of the jet energy absorbed by the Barrel Presampler</td>
</tr>
<tr>
<td>EMBFrac</td>
<td>The fraction of the jet energy absorbed by the Electromagnetic Barrel (EMB) calorimeter</td>
</tr>
<tr>
<td>EMBXFrac</td>
<td>The fraction of the jet energy absorbed in the X = 1, 2, 3 layer of the EMB calorimeter</td>
</tr>
<tr>
<td>Mass</td>
<td>The mass of the jet calculated by summing the 4-vectors of the constituents then using $m^2 = E^2 - P^2$</td>
</tr>
<tr>
<td>TILEBXFrac</td>
<td>The fraction of the energy absorbed in the X = 0, 1, 2 layer of the Tile Barrel (TILEB) calorimeter</td>
</tr>
<tr>
<td>EMBXOverEMB</td>
<td>The energy absorbed in the X = 1, 2, 3 layer of the EMB calorimeter divided by the energy absorbed in all layers of the EMB calorimeter</td>
</tr>
<tr>
<td>TILEXOverTILE</td>
<td>Energy absorbed in the X = 0, 1, 2 layer of the TILE calorimeter divided by the energy absorbed in the TILE</td>
</tr>
<tr>
<td>MostELayer</td>
<td>The calorimeter layer which absorbed the most energy</td>
</tr>
<tr>
<td>MostELayerFrac</td>
<td>The fraction of the energy absorbed in the layer which absorbed the most energy</td>
</tr>
<tr>
<td>MostEperX0</td>
<td>The energy deposit per radiation length in the layer which had the highest energy deposit per radiation length</td>
</tr>
<tr>
<td>EndLayer</td>
<td>The calorimeter layer by which 95% of the calorimeter energy is absorbed</td>
</tr>
<tr>
<td>NumTowers</td>
<td>The number of calorimeter towers in which the jet deposited energy</td>
</tr>
</tbody>
</table>

Table 6.1: Jet variables and a brief description
Figure 6.3: The NTrk500 (a), Width (b), TILEB0Frac (c), NumTowers (d), TrackPt1Frac (e), and EMBFrac (f) distributions. Data are shown in black, Pythia is in red and Herwig++ is in blue for the reconstructed $p_T$ range 105-125 GeV.
In the following figures the high response jets are labelled as “signal” and the low response jets as “background”. The signal and background distributions for NTrk500 are shown in figure 6.4. Signal jets favour a low NTrk500 as they tend to have a high $\pi^0$ content from the parton shower, so there are fewer charged tracks. Signal jets tend to have a low

![Figure 6.4: NTrk500 signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right).](image)

TrackPt500Frac, shown in figure 6.5, for the same reason.

![Figure 6.5: TrackPt500Frac signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right).](image)

The AVGTrackPt500Frac shown in fig 6.6 was found to have little ability in distinguishing between signal and background. This is probably because the TrackPtFrac and the number of tracks are correlated. If a jet has a small number of tracks, the tracks most likely
carry a small fraction of the momentum of the jet and if a jet has a large number of tracks the tracks most likely carry a large fraction of the jet momentum.

Figure 6.6: AVGTrackPt1Frac signal and background distributions for \( p_T \) bins of 105-125 GeV (left) and 500-600 GeV (right).

TrackWidth shown in 6.7 was found to give some separation of signal and background distributions. Signal jets contain more \( \pi^0 \)'s and thus fewer tracks, giving signal jets a smaller TrackWidth.

Figure 6.7: TrackWidth signal and background distributions for \( p_T \) bins of 105-125 GeV (left) and 500-600 GeV (right).

Width was found to have better separation ability than TrackWidth. Width is built using energy deposits within the calorimeter, giving some indication of how the jet showered in the calorimeter. Signal jets favour a small Width as they have a larger electromagnetic component meaning most of the energy is deposited close to the jet axis.
EMBFrac, shown in Figure 6.9, is very good at separating signal and background jets. Signal jets have a large electromagnetic component meaning most of their energy is absorbed in the EMB calorimeter.

![Graphs showing signal and background distributions for different pT bins.](image)

**Figure 6.8:** Width signal and background distributions for pT bins of 105-125 GeV (left) and 500-600 GeV (right).

EMB1Frac, shown in Figure 6.10, gives a small separation between signal and background. The full electromagnetic shower is not usually contained within the EMB1 layer, which is relatively thin so the fraction of energy a jet deposits in this layer can be similar for high and low response jets.

EMB2Frac is very good at separating signal and background distributions. EMB2 is the largest layer in the electromagnetic calorimeter and most of the electromagnetic component...
of a jet is absorbed in this layer. Signal jets have a large electromagnetic component meaning most of their energy is absorbed in this layer. The EMB2Frac has a very similar distribution of signal and background jets as the EMBFrac shown in fig 6.9.

Electromagnetic energy is mostly absorbed in the second layer of the EMB with some leaking into EMB3. Low response jets penetrate deeper into the calorimeter as they have a larger hadronic component and therefore tend to have a slightly higher EMB3Frac than signal jets shown in figure 6.11.

Mass gives better separation at higher $p_T$ because jets are able to take on a larger range of mass values. At lower energy, where the range of mass values is limited, mass is unable
to give much separation between signal and background distributions. High mass particles from a hard scattering have a smaller Lorentz boost in the direction of scattering than low mass particles. This means their decay products are also less boosted in this direction so when they reach the detector they will be more spread out than decay products from low mass particles. For this reason, high mass particles have a larger Width than low mass particles. As was seen before, high response particles have a low Width and so high response particles have a low mass. Figures 6.23 and 6.24 show the two are highly correlated. It is possible that a high mass particle will have a larger fraction of particles which lie outside the jet definition. Particles decaying from high mass particles, being less boosted in the direction of scattering, are able to travel away from the center of the jet, making them more likely to hit the detector outside of the jet definition.

![Mass signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right).](image)

Figure 6.12: Mass signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right).

Signal jets tend to have a low TILEB0Frac, TILEB1Frac and TILEB2Frac as most of their energy has already been absorbed in the EMB. TILEB0Frac, shown in figure 6.13 gives the best separation between signal and background distributions of the three variables. This is because TILEB0 is the first layer in the TILE. As a jet with high response travels deeper into the TILE calorimeter it has less energy to deposit, causing the fraction of energy deposited to go down.

EMB1OverEMB, EMB2OverEMB and EMB3OverEMB are all somewhat anti-correlated as energy going into one layer will take away from energy going into a different layer. The correlation is shown in figures 6.23 and 6.24. EMB2OverEMB and EMB3OverEMB are less anti-correlated because its possible for a large fraction of energy to be deposited in EMB1, making EMB2OverEMB and EMB3OverEMB both small. Signal jets tend to deposit most of their energy in EMB2 as it is the largest EMB layer and signal jets have a large electromagnetic component. This means that signal jets have a slightly higher EMB2OverEMB
Figure 6.13: TILEB0Frac signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right).

(figure 6.14) and a slightly lower EMB1OverEMB and EMB3OverEMB than background jets.

Figure 6.14: EMB2OverEMB signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right).

TILEB0OverTILE, TILEB1OverTILE and TILEB2OverTILE are all anti-correlated, as demonstrated in figures 6.23 and 6.24, for the same reason as the EMBXOverEMB variables. The signal and background distributions for all three variables are very similar because the hadronic component of both signal and background jets interact with the TILE calorimeter in the same way. Figure 6.15 shows the signal and background distributions for TILEB1OverTILE.
Figure 6.15: TILEB1OverTILE signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right).

MostELayer does not give any separation between signal and background jets. The majority of jets deposit the largest amount of energy in EMB2. The electromagnetic component of both signal and background jets will deposit the majority of their energy in this layer. Bremsstrahlung from charged hadrons will also be deposited in this layer more than any other. This leads to very similar distributions for both signal and background shown in figure 6.16. As the layer in which a jet deposits most of its energy is EMB2, MostELayer-Frac is highly correlated with EMB2Frac (see figures 6.23 and 6.24) and has similar signal and background distributions.

MostEperX0 provides some separation of signal and background distributions. This is expected as a high response jet has a high energy density.

EndLayer provides good separation between signal and background jets. Signal jets, containing a large electromagnetic component, tend to deposit 95% of their energy before reaching the TILE. Background jets usually do not deposit 95% of their energy before the TILE as they have a much larger hadronic component than signal jets.

NumTowers signal and background distributions are fairly well separated because signal jets tend to be smaller in size than background jets. The NumTowers distribution is shown in figure 6.19.

Five quality measures were used to rank the variables according to how well they distinguish between high and low response jets; an integrated probability, an integrated signal-weighted probability, an integrated event-weighted signal/background ratio, the mean asymmetry and a Chi2 test between the high response and low response distributions. In each bin of a variable the number of high response jets was taken as signal, $s$, and the number of low response jets was taken as background, $b$. The integrated probability, $P$ is then given
Figure 6.16: MostELayer signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right). The calorimeter is labelled from the inside-out, with the Presampler = 0 and TILEB2 = 7. EMB2 is labelled 3. Label 8 corresponds to punch-through energy that hits the muon detectors.

Figure 6.17: MostEperX0 signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right).

by,

$$ P = \int \frac{s}{s + b}. $$

(6.2)

The integrated probability, while giving a measure of how well the signal and background distributions are separated, can be influenced by bins which have a low number of events. For instance a bin with one signal event and no background events contributes a large amount to the integrated probability. A large number of bins similar to this can give a
Figure 6.18: EndLayer signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right). The calorimeter is labelled from the inside-out, with the Presampler = 0 and TILEB2 = 7. EMB2 is labelled 3. Label 8 corresponds to punch-through energy that hits the muon detectors.

Figure 6.19: NumTowers signal and background distributions for $p_T$ bins of 105-125 GeV (left) and 500-600 GeV (right).

variable a high integrated probability but the variable might not be very good at separating signal and background distributions where the signal is large.

To test for this, a signal-weighted integrated probability (SWP) given by,

$$SWP = \int \frac{s^2}{s+b},$$

$$\text{(6.3)}$$
and the integrated event-weighted signal/background ratio (EWSB) given by,

\[
WSB = \int \frac{(s + b) \times s}{b}.
\]  

(6.4)

are also used. These quality measures take into account the number of events in a bin and avoid biasing variables with regions that do not contribute to the overall signal. The probability, SWP and event WSB in the range 105-125 GeV are shown for EMBFrac in figures 6.20, 6.21 and 6.22 respectively. It is clear in this example that low EMBFrac is incorrectly identified as a good region by the simple probability, while the other quality measures remedy the situation.

The mean asymmetry, \( M_A \), of each variable was calculated using the mean of both the signal, \( M_s \) and background, \( M_b \) distributions as,

\[
M_A = \frac{M_s - M_b}{M_s + M_b}.
\]  

(6.5)

The mean asymmetry is included because it gives a direct measure of the separation of the signal and background distributions. A weakness of this quality measure is that it does not take into account the width of the distributions. A variable that has a large mean asymmetry but has very broad signal and background distributions is not that useful.

The Chi2 test is included because it gives a measure of how similar the signal and background distributions are. A weakness of this quality measure is that it does not work well for variables with a small number of bins like the number of tracks or any of the “Layer” variables.

![Figure 6.20: Probability distribution for EMBFrac in the range 105-125 GeV](Image)
After the calculations are performed, the variables are ranked according to each quality measure. The ranks are then combined to get an overall rating.

Variables which are highly correlated give no new information and therefore it is not very useful to use a large number of correlated variables. A study of the correlations between jet variables was undertaken. Figures 6.23 and 6.24 show the correlation factors for jet variables in Pythia for 105-125 GeV and 500-600 GeV respectively. The correlation factor is determined by dividing the covariance of two variables by the product of their standard
deviations. This results in values ranging from -1 for purely anti-correlated variables to 1 for purely correlated variables.

There are some variable correlations for which the reason is obvious: EMBXFrac will be correlated with EMBXOverEMB; EMBFrac will be anti-correlated with TILEB0Frac, TILEB1Frac and TILEB2Frac; and TrackPtXFrac will be correlated with AVGTrackPtXFrac. A strong correlation exists between EMB1Frac and MostEperX0 but the reason is slightly more subtle. EMB1 is a relatively thin (in terms of radiation lengths) calorimeter layer which often receives a significant fraction of a jet’s energy as it is one of the first layers the jet encounters. Therefore, the higher the EMB1Frac, the more energy is deposited in a relatively thin layer, giving this layer a high MostEperX0.

TILEB0Frac and MostELayerFrac have a strong anti-correlation due to the fact that the MostELayer is usually EMB2. As EMB2Frac is anti-correlated with TILEB0Frac, MostELayerFrac will be anti-correlated with TILEB0Frac. Appendix H shows the correlation factors for the different $p_T$ ranges between 25-2500 GeV for Pythia and Herwig++. Correlation factors for data are also shown for $p_T$ ranges between 100-800 GeV. Fair agreement is seen between data and simulation.

6.2 Correcting the Reference Jet Energy

In order to use variable information to determine the response of a jet, a likelihood function, $L$, is considered. $L$ is given by,

$$L = \prod_{V} P_{V(x)}$$

(6.6)

where $P_{V(x)}$ is the probability that given the jet has value $x$ for variable $V$, the jet has a large response.

The overall ranks of the variables were determined by adding the ranks of the individual quality measures together with a lower sum indicating a better variable. These ranks were used as a guide when selecting variables as there were too many variables (approximately 35 variables in 17 different momentum ranges) to do select variables entirely “by eye”. Using these ranks as a guide and the correlation factors, 6 variables were selected by hand for a likelihood function in each momentum range. The first variable selected was the top ranked variable after which variables were discarded if the magnitude of their correlation factor with another already selected variable was greater than ±0.4. Two additional likelihood functions for each momentum range were made; one using 14 variables with the magnitude of correlation factors less than 0.8 and the other using 9 variables with the magnitude of correlation factors less than 0.6 (Intermediate). These additional likelihood functions are used to test systematic uncertainties on the method. The two likelihood variations above along with the three likelihood variations using different high response and low response definitions, make for a total of five additional likelihood variations in each $p_T$ range. The 6 (1 nominal and 5 variations) likelihoods are calculated in both Pythia and Herwig++. The
Figure 6.23: Correlations between jet variables in Pythia for reconstructed $p_T$ from 105-125 GeV
Figure 6.24: Correlations between jet variables in Pythia for reconstructed $p_T$ from 500-600 GeV
variables the likelihood functions use are selected using Pythia data. The same variables are used in Herwig++. The selected variables in the various $p_T$ ranges are given in Appendix B.

The MCResponse versus -log(L) distribution for 105-125 GeV is shown in Fig. 6.25, where a definite correlation can be seen. -Log(L) bins were made to test the relationship

![MCResponse vs. Loglikelihood](image)

Figure 6.25: MCResponse vs -log(L) in the momentum range 105-125 GeV. Colour indicates the number of jets between likelihood and MCResponse. -Log(L) bins were defined such that there were at least 1500 Pythia events in the corresponding two-dimensional - log(L)/$p_T$ bin. The response distribution for each likelihood bin in each reconstructed momentum range was fit using a Gaussian function. The distribution was then refit in the region ± 2 standard deviations around the mean of the first fit function. The resulting fits for the momentum range 105-125 GeV, along with the response when integrating over likelihood, are shown in Figure 6.26. This demonstrates that binning in likelihood results in a much narrower response distribution, as desired. It should be noted that a lower -log(L) corresponds to a higher response jet.

The mean of each fit is used as the response correction for jets in the corresponding in the corresponding two-dimensional -log(L)/momentum bin. This correction is referred to as the “$\alpha$ correction,” for historical reasons. An estimate of the particle level jet energy is given by the reconstructed jet energy divided by the $\alpha$ correction. Figure 6.27 shows the $\alpha$ correction as a function of -log(L) for the momentum ranges 105-125 GeV and 500-600 GeV.
The closure of the alpha correction, $\alpha_c$ gives a measure of how close the $\alpha$ corrected momentum, $p_T^{\alpha}$ is to the truth momentum in Monte Carlo simulations, $p_T^{Truth}$. $\alpha_c$ is calculated using,

$$\alpha_c = \frac{p_T^{\alpha} - p_T^{Truth}}{p_T^{Truth}},$$  \hspace{1cm} (6.7)
Figure 6.27: The mean response fit vs likelihood for jets with momentum 105-125 GeV (left) and 500-600 GeV (right) for each momentum range where $p_T^\alpha = p_T^{\text{rec}} / \alpha$. The $\alpha$ distributions are then fit twice using a Gaussian function in the same manner as the response distributions. The offset of the mean from zero of this distribution is used as a residual correction to ensure that the $\alpha$ corrected reference jet has the same momentum as the particle level jet. The standard deviation, $\sigma$, of this distribution is taken as the uncertainty on the reference jet. $\alpha_c$ as a function of momentum in Pythia is shown in figure 6.28 and the standard deviation is shown in figure 6.29.

Figure 6.28: The alpha closure as a function of momentum in Pythia

Using the $\alpha$ correction, reference jets have now been created and are ready for use in the MPF technique.
Figure 6.29: The standard deviation of the alpha closure as a function of momentum in Pythia
Chapter 7

Determining the Jet Response

7.1 Data Samples

This thesis makes use of 20.8 fb\(^{-1}\) of \(\sqrt{s} = 8\) TeV data recorded between April 4th and December 12th 2012. Dijet results derived using these data are compared to the same Monte Carlo generators, Pythia8 and Herwig+++, as were used for the previously discussed jet variable study. \(\gamma +\) jet results are compared to events generated independently by Pythia8 and Herwig+++

7.2 Selection Criteria

7.2.1 photon selection

Photon candidates are selected using the sliding window approach discussed in section 4.1.1. For this analysis photons were required to have at least 25 GeV in energy and a pseudorapidity of less than 1.37 to avoid the crack region of the EM calorimeter. A variety of cuts are made to ensure the energy deposits in the calorimeter are consistent with an electromagnetic shower. One such cut is placed on the energy deposited within the EM calorimeter versus the hadronic calorimeter. This cut uses the fact that hadronic showers have a much larger shower depth than electromagnetic showers. A similar cut is placed on the fraction of the energy within the third layer of the EM calorimeter. These two cuts are highly correlated.

As was mentioned previously, hadronic showers are much wider than electromagnetic showers. To take advantage of this, a cut is placed on the size of the energy deposit in \(\eta\) (but not \(\phi\)) space in both the strip and second EM calorimeter layer. A second isolation criterion is applied where 3x7 and 7x7 sliding cluster windows are both centred on the energy deposit and a requirement is placed on the ratio of the energy contained within the clusters. To reduce hadronic fakes which are dominated by \(\pi^0 \rightarrow \gamma \gamma\) a limit is set on the ratio of energy in the two highest energy strips in the EM calorimeter. Events are rejected
where this ratio is above a certain threshold which depends on energy. This threshold is at least 0.24 [20].

For photons with energy 25-45 GeV, 45-65 GeV, 65-85 GeV and ≥ 85 GeV a cut of 0.5 GeV, 1 GeV, 2 GeV, and 3 GeV is placed on the energy within a cone of radius R = 0.4 around the photon, excluding the photon energy. This cut helps to reduce jets faking photons. A cut is placed on the ratio of cluster energy to track momentum for tracks coming from photons. For a photon with one track the E/P ratio must be below 2 while for fully converted photons the E/P ratio must be between 0.5 and 1.5 for the two tracks. Events are vetoed if they do not pass this criterion.

Photons in data must also pass the photon trigger. The six triggers and their ranges are listed in table 7.1. The 99% efficiency level is the momentum beyond which 99% of photons pass the trigger.

7.2.2 Jet Selection

Jets were reconstructed using 4-2-0 topoclusters and the anti-$k_T$ algorithm with a size parameter of R = 0.6. Selection criteria were based on the in-situ-corrected jet momentum. Both the leading and subleading jets in a dijet event were required to have $p_T \geq 8$ GeV and $\eta \leq 0.8$. The leading and subleading jets were required to have a jet vertex fraction greater than 0.25. The anti-$k_T$ algorithm does not distinguish between sources of energy deposits, meaning that it is possible that the algorithm builds a jet from a lepton. To avoid these fake jets, any jet within $\Delta R = 0.3$ of a reconstructed photon is vetoed. Events where the leading or subleading jet was affected by noise bursts in the calorimeter (bad jets) were vetoed. Events where the leading or subleading jet fell in problematic regions of the detector (ugly jets) were also vetoed. To avoid energy contributions from nearby jets, events were thrown out if there was a jet with over 10 GeV of energy within R = 1.2 of a leading and subleading jet.

One jet in a dijet event is required to pass one of the jet triggers listed in table 7.2. $p_T^{avg}$ is the average $p_T$ of the leading and subleading jets. A single jet trigger was applied in each calibrated scale $p_T^{avg}$ range, as suggested by the ATLAS JetEtMiss group [27]. It should be noted that EF-j15-a4tchad was the lowest $p_T$ trigger with a 99% efficiency level.

<table>
<thead>
<tr>
<th>Trigger Name</th>
<th>99% efficiency level (GeV)</th>
<th>$p_T$ range applied (GeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF-g20-loose</td>
<td>20</td>
<td>25-45</td>
</tr>
<tr>
<td>EF-g40-loose</td>
<td>40</td>
<td>45-65</td>
</tr>
<tr>
<td>EF-g60-loose</td>
<td>60</td>
<td>65-85</td>
</tr>
<tr>
<td>EF-g80-loose</td>
<td>80</td>
<td>85-105</td>
</tr>
<tr>
<td>EF-g100-loose</td>
<td>100</td>
<td>105-125</td>
</tr>
<tr>
<td>EF-g120-loose</td>
<td>120</td>
<td>125-∞</td>
</tr>
</tbody>
</table>

Table 7.1: Photon Triggers
<table>
<thead>
<tr>
<th>Trigger Name</th>
<th>99 % efficiency level (GeV)</th>
<th>$p_T^{avg}$ range applied (GeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF-j15-a4tchad</td>
<td>33</td>
<td>25-55</td>
</tr>
<tr>
<td>EF-j25-a4tchad</td>
<td>49</td>
<td>55-70</td>
</tr>
<tr>
<td>EF-j35-a4tchad</td>
<td>63</td>
<td>70-100</td>
</tr>
<tr>
<td>EF-j55-a4tchad</td>
<td>85</td>
<td>100-135</td>
</tr>
<tr>
<td>EF-j80-a4tchad</td>
<td>126</td>
<td>135-180</td>
</tr>
<tr>
<td>EF-j110-a4tchad</td>
<td>150</td>
<td>180-200</td>
</tr>
<tr>
<td>EF-j145-a4tchad</td>
<td>183</td>
<td>200-240</td>
</tr>
<tr>
<td>EF-j180-a4tchad</td>
<td>220</td>
<td>240-300</td>
</tr>
<tr>
<td>EF-j220-a4tchad</td>
<td>262</td>
<td>300-480</td>
</tr>
<tr>
<td>EF-j360-a4tchad</td>
<td>403</td>
<td>480-inf</td>
</tr>
</tbody>
</table>

Table 7.2: Jet Triggers for calibrated scale $p_T^{avg}$

at 33 GeV while the lowest $p_T$ bin is from 25-45 GeV. This means that in this bin some events below 33 GeV are lost, with a larger fraction of events lost at lower $p_T$. This can be seen in Figures 7.1 and 7.2.

### 7.2.3 Event Selection

In the derivation of the MPF response it was assumed that the transverse momentum of the two outgoing particles balance each other but this is not always true. It is possible that an out-going parton radiates a parton, known as final state radiation (FSR) such that the original out-going particles no longer satisfy the assumption. In addition to FSR there is also initial state radiation (ISR) where an incoming parton radiates a parton, which also ruins the assumption. To reduce the effects of ISR and FSR the reference object and probe jet are required to be back-to-back i.e. have a $\Delta \phi$ between them greater than 2.9. Furthermore a minimum cut of 12 GeV or $0.2 \times p_T^{leading\,jet}$ (whichever is greater) is placed on the subleading jet in $\gamma +$ jet and third-leading jet in dijet events if the third-leading jet has a jet vertex fraction greater than 0.75.

The leading and subleading jet $p_T$ are shown in figures 7.1 and 7.2 respectively. The shapes of the distributions in data and simulation are in good agreement above 105 GeV while below 105 GeV there is a deficit in data events. In the 25-45 GeV bin this deficit is due to the trigger not being fully efficient. Further study will be needed to determine the source of the discrepancy between data and simulations between 45 and 105 GeV. However, this discrepancy should not affect the results presented in this thesis, which do not depend on the dijet cross section.

The fraction of the leading jet $p_T$ carried by the third-leading jet and the $\Delta \phi$ between the leading and subleading jets are shown in figures 7.3 and 7.4 for the $p_T$ range 160-210 GeV. Fair agreement is seen between data and simulation other than some fluctuations in the third leading jet distribution in the simulations. Appendix C shows the third leading jet $p_T$ and $\Delta \phi$ distributions for the other $p_T$ ranges considered.
Figure 7.1: The leading jet $p_T$ in dijet events at calibrated scale for data (black), Pythia (red) and Herwig++ (blue).

Figure 7.2: The subleading jet $p_T$ in dijet events at calibrated scale in data (black), Pythia (red) and Herwig++ (blue).

In photon + jet events the low $p_T$ bins (less than 125 GeV) are chosen to be the same ranges where the triggers are applied. This strategy results in very good agreement between data and simulation for the leading jet, subleading jet, and $\Delta\phi$(jet, ref) distributions [5, 20].
Figure 7.3: The fraction of the leading jet $p_T$ carried by the third leading jet in data (black), Pythia (red) and Herwig++ (blue) for the $p_T$ range 400-500 GeV.

Figure 7.4: The $\Delta \phi$ between the leading and subleading jets in data (black), Pythia (red/needs to be added) and Herwig++ (blue) for the $p_T$ range 400-500 GeV.
7.3 Measuring the Jet Response

The jet response must be measured in bins of $p_T$ to account for the dependence on particle level jet energy. The transverse momentum of the photon (alpha corrected jet) can be used to give a measure of the particle level jet $p_T$ in photon + jet (dijet) events. Within each $p_T$ bin the response is fit using a Poisson distribution and then refit between the mean $\pm 2\sigma$ as described in [5]. The bin edges are 25 GeV, 45 GeV, 65 GeV, 85 GeV, 105 GeV, 125 GeV, 160 GeV, 210 GeV, 260 GeV, 310 GeV, 400 GeV, 500 GeV, 600 GeV, and 800 GeV for both dijet and $\gamma +$ jet events. The dijet response distribution and the fit from data in the $p_T$ range 160-210 GeV are shown in figure 7.5.

![Figure 7.5: The dijet response distribution from data fit with a Poisson function for the $p_T$ range 160-210 GeV. The mean listed in this plot indicates the difference between 1 and the measured response.](image)

7.4 Systematic Uncertainties

Differences in the measured MPF response in data and Monte Carlo simulation may be due to mismodeling of the detector, mismodeling of initial and final state radiation, or mismodeling of the reference object. A selection of parameters are varied and the variation in the response is used as a measure of its uncertainty. Effects studied in this thesis are,

- Initial and final state radiation, varied by changing the conditions on the second leading jet (or third leading in dijet) and $\Delta\phi(jet, \text{ref})$
• Signal and background definitions in the alpha correction
• The choice of variables used in the likelihood function
• Uncertainty on the reference object energy scale from the alpha closure width

7.4.1 ISR/FSR

To reduce the effect of ISR/FSR on the MPF response additional jets with large momentum are vetoed. The sub-leading jet cut in $\gamma + \text{jet}$ and the third-leading jet cut in dijet are varied between a ‘loose’ value of $0.3 \times p_T^{\text{leadJet}}$ or a ‘tight’ value of $0.1 \times p_T^{\text{leadJet}}$ or 10 GeV. The $\Delta \phi(\text{jet, ref})$ is varied from a looser cut of 2.8 to a tighter cut of 3.0 in both $\gamma + \text{jet}$ and dijet events. The result of these variations in dijet data, Pythia dijet, $\gamma + \text{jet}$ data and Pythia $\gamma + \text{jet}$ are shown in figures 7.6, 7.7, 7.8 and 7.9, respectively. The relative response difference, $\Delta_{\text{rel}}$, between the nominal and a varied selection is given by,

$$\Delta_{\text{rel}} = \frac{R_{\text{var}} - R_{\text{Nom}}}{R_{\text{Nom}}}$$  (7.1)

where $R_{\text{var}}$ is the response of the varied selection and $R_{\text{Nom}}$ is the response of the nominal selection. The relative differences between the nominal and varied ISR/FSR selection are below 5% across all momentum ranges in both data and simulation with this difference decreasing as $p_T$ increases. A similar effect is seen in $\gamma + \text{jet}$ events for both data and simulation but with a maximum difference of 2.5%. The envelopes of the $\Delta \phi$ and $J2/J3$ variations are added in quadrature in dijet and $\gamma + \text{jet}$ events and taken as a systematic uncertainty on the MPF response in the respective sample. The systematic uncertainties are asymmetric so the final MPF response uncertainty will be asymmetric as well.

7.4.2 Signal/Background Definitions

The definition of signal and background jets in the log-likelihood function changes the shape of the log-likelihood distribution and therefore the log-likelihood value assigned to a jet. The log-likelihood bins used must change with each definition to ensure there are enough jets in each log-likelihood bin. The alpha correction for the new log-likelihood bins must be changed to reflect the response of the jets in the new log-likelihood bins. This results in slightly different alpha corrections for each log-likelihood function.

This means that the reference jet for each log-likelihood function has a slightly different $p_T$ which directly affects the response measurement of the probe jet. The relative differences in response when using the nominal 10% signal-70% background and the varied 10% signal-50% background, 20% signal-50% background and 20% signal-70% background likelihood functions are shown for data in figure 7.10 and for Pythia in figure 7.11. The relative difference in the MPF response when using the nominal and varied log-likelihood
Figure 7.6: Relative response difference between the nominal and varied ISR/FSR cuts in dijet data events

Figure 7.7: Relative response difference between the nominal and varied ISR/FSR cuts in dijet PYTHIA events
Figure 7.8: Relative response difference between the nominal and varied ISR/FSR cuts in $\gamma + \text{jet Data events}$

Figure 7.9: Relative response difference between the nominal and varied ISR/FSR cuts in Pythia $\gamma + \text{jet events}$
functions is fairly large and can be as high as 0.12. Only the largest difference, both positive and negative, is added in quadrature to the asymmetric systematic uncertainties, with positive differences added to the upper uncertainty and negative differences to the lower uncertainty. Figure 7.10 shows that the nominal log-likelihood function usually gives a lower MPF response than the varied log-likelihoods. This results in a large uncertainty above the measured dijet MPF response.

Figure 7.10: Relative response difference between the nominal 10% signal-70% background likelihood function and the varied 10% signal-50% background, 20% signal-50% background and 20% signal-70% background likelihood functions in data dijet events

7.4.3 Correlation Effects

To measure the effect of using correlated variables in the log-likelihood function on the jet MPF response two alternative log-likelihood functions were defined. One used 14 variables that could be highly correlated (correlation factors ≤ 0.8) and the other used 9 variables that could have moderate correlations (correlation factors ≤ 0.6). The nominal log-likelihood function used 6 variables that were relatively uncorrelated (correlation factors ≤ 0.4). The nominal 10% signal-70% background definitions were used for all three log-likelihood functions. As different variables were used in each log-likelihood function, each log-likelihood function gives a different alpha correction for the reference jet. This results in a different measured MPF response for each log-likelihood function. The relative differences of the response between the nominal and varied log-likelihood functions are shown for data in figure 7.12 and for Pythia in figure 7.13. The relative response differences between the nominal and
Figure 7.11: Relative response difference between the nominal 10% signal-20% background likelihood function and the varied 10% signal-50% background, 20% signal-50% background and 20% signal-70% background likelihood functions in Pythia dijet events.

varied log-likelihoods have magnitudes less than 8% across all $p_T$ ranges. These differences are slightly less than the differences from varying the signal and background definitions but are still a large component of the dijet response uncertainty. The largest relative differences, both positive and negative, were added in quadrature to the asymmetric dijet uncertainties as before.

7.4.4 Alpha Closure $\sigma$

The $\sigma$ of the alpha closure gives a measure of the reference jet energy uncertainty. This error creates an uncertainty in the MPF response as the reference object could have a true energy that is higher or lower than the alpha corrected jet energy. To account for this uncertainty the $\sigma$ of the alpha closure is added in quadrature to the systematic uncertainty. As was seen in figure 6.29 this uncertainty ranges from 14% at low energy to about 4% as high energy.

7.4.5 Photon + Jet Systematic Uncertainty

Sources of systematic uncertainty on the response in $\gamma +$ jet events include the photon energy scale, the photon purity and pile-up. Any uncertainty in the photon energy scale directly affects the response uncertainty as the photon is used to infer the true energy of
Figure 7.12: The relative response differences when using the nominal 14 variable log-likelihood functions and the varied 6 and 9 variable log-likelihood functions in data dijet events.

Figure 7.13: The relative response differences when using the nominal 14 variable log-likelihood functions and the varied 6 and 9 variable log-likelihood functions in Pythia events.

It has been shown [5, 6, 20] that variations in the photon energy scale affect the MPF response much less than 1%.
In the lowest $p_T$ ranges $\gamma +$ jet events suffer from contamination of dijet events where a jet has passed all of the photon criteria. This ‘fake’ photon has a lower response than a real photon since it is really a jet. Using the fake as a reference object results in a higher MPF response because the measured energy of the reference is lower than its true energy. The purity is roughly the fraction of photons in a $p_T$ bin which are truly photons and not a jet faking a photon. Uncertainties in the purity result in a 0.5 % uncertainty at low $p_T$ and become negligible at higher $p_T$ [5].

As jets are much larger in size than photons, jets contain a larger amount of pile-up and therefore are more susceptible to changes in pile-up conditions. In Ref. [5] the leading jet vertex fraction was varied to allow more or less pile-up jets into $\gamma +$ jet events. The number of primary vertices (NPV) and the average number of interactions per bunch crossing $\langle \mu \rangle$ were also varied. After applying the pile-up corrections previously discussed it was found that the overall effect of pile-up on the response was about 0.2%. It should be noted that in a dijet event the two jets contain roughly equal amounts of pile-up so that the effect of pile-up on the response in dijet events is negligible.

To take these sources of uncertainty into account in the $\gamma +$ jet response a flat 1% uncertainty is added to the $\gamma +$ jet response.

### 7.5 Results

In dijet events the largest uncertainty from the $\Delta \phi$ variation, third jet variation, signal-background variation, jet variable variation and the reference jet energy uncertainties have been added in quadrature. Several components of the dijet MPF response uncertainty were relatively large and range from 0.02-0.17. Figure 7.14 shows that below 105 GeV the dijet response uncertainty is large (0.1-0.2). This is because the dijet response uncertainty in this region is dominated by the uncertainty on the tag jet $p_T$ which comes from the resolution of the alpha closure shown in Figure 6.29. Above 105 GeV the tag jet $p_T$ uncertainty becomes comparable to the sources of uncertainty from varying the definitions of the likelihood function and range from 0.04-0.1. The size of the uncertainties in simulation is nearly the same size as the uncertainties in data.

The dijet MPF response ranges from 0.5 at low $p_T$ to 0.75 at high $p_T$ as shown in Figure 7.14. The dijet response in data agrees with the Pythia dijet response within 0.02 across all momentum ranges except between 65-85 GeV where this difference is 0.04. The Herwig++ dijet response is 0.02-0.08 higher than both the data and Pythia results but still agrees within uncertainties. The small disagreement between data and Monte Carlo simulations is most likely due to the alpha corrected jet energy being slightly different than the particle level jet energy in data. In Monte Carlo simulation the residual correction ensured that the alpha corrected jet had the same energy as the particle level jet. It may be possible to determine a residual correction for data by using $\gamma +$ jet events and comparing the alpha
corrected jet energy to the photon energy but doing this study is beyond the scope and time constraints of the current project.

Figure 7.14: The dijet MPF response after using the alpha-corrected jet as a reference object. All sources of uncertainty are included. Data and simulation markers have been offset by 3 GeV from each other for ease of comparison. Data are shown in black, Pythia in red and Herwig++ in blue.

In $\gamma$ + jet events the sources of uncertainty from the $\Delta\phi$ variation, the second jet variation and a flat 1% uncertainty were added in quadrature. The flat 1% uncertainty covers the combined uncertainties from the photon purity, photon energy scale, and pile-up. Figure 7.15 shows that the $\gamma$ + jet response in data and Pythia agree within 0.02 for most momentum ranges. The response in $\gamma$ + jet events predicted by Herwig++ is 0.02-0.04 higher than the response in data. Data uncertainties and simulation uncertainties are both in the 0.01-0.03 range. Below 85 GeV simulations predict a 0.01-0.03 lower response than data while above 85 GeV simulations predict a 0.01-0.04 higher response. The exception to this is Herwig++ in the range 25-45 GeV which predicts a slightly higher response than what is seen in data. The size of the uncertainties in $\gamma$ + jet are much smaller than dijet because the momentum of the reference object is known with much greater accuracy.
Figure 7.15: The $\gamma + \text{jet}$ MPF Response. All sources of uncertainty are included. Data and simulation markers have been offset by 3 GeV from each other for ease of comparison. Data are shown in black, Pythia in red and Herwig++ in blue.
Chapter 8

Calculating the Quark and Gluon Jet Responses

The $\gamma +$ jet and dijet responses have now been determined in data and simulations and it is possible to calculate quark and gluon responses using equations 5.20 and 5.21. These equations also require the parton fractions for jets in dijet and $\gamma +$ jet events shown in figures 8.1 and 8.2. Some disagreement is seen between the quark and gluon fractions predicted by Pythia versus Herwig++ in both dijet and $\gamma +$ jet events. This disagreement is used as a systematic uncertainty in the final quark and gluon responses. The disagreement for the c-quark fraction is small and is ignored in the calculation of the quark and gluon response uncertainties.

The c-quark MPF response shown in Fig. 8.3 is required to calculate the quark and gluon responses and is estimated from Monte Carlo simulation. Small differences between Pythia and Herwig++ are seen in the predicted c quark response. These differences are included as a systematic uncertainty in the calculated quark and gluon responses.

8.1 Calculating Uncertainties

To properly incorporate uncertainties in the measured results into the calculated quark and gluon response uncertainties the way the quark and gluon responses change as a function of their inputs needs to be determined. The component of the uncertainty on the quark response, $R_q$, from the uncertainty in the dijet quark fraction $f_q^{DJ}$ is given as

$$\Delta R_q\left(f_q^{DJ}\right) = \frac{dR_q}{df_q^{DJ}} \times \Delta f_q^{DJ}. \hspace{1cm} (8.1)$$

The derivatives of the quark and gluon responses with respect to each of their inputs are calculated in appendix A.

The sources of uncertainty on the quark and gluon response are:
Figure 8.1: The jet parton fractions in dijet events. The solid lines are Pythia and the dashed lines are Herwig++. The gluon fraction is shown in blue, the light-quark fraction in red and the c-quark fraction in black.

Figure 8.2: The jet parton fractions in $\gamma +$ jet events. Solid lines are Pythia and the dashed lines are Herwig++. The gluon fraction is shown in blue, the light-quark fraction in red and the c-quark fraction in black.
The $c$-quark fraction $f_c$ is assumed to be known well enough to ignore in calculation of the uncertainties of the quark and gluon responses. Figures 8.1 and 8.2 show that the $f_c$ uncertainty is quite small, which further justifies ignoring its uncertainty. Possible correlations must be taken into account when calculating uncertainties. The relationship, $f_q + f_g + f_c = 1$ is used to remove $f_g$ from equations 5.20 and 5.21 as $f_g$ is 100% anti-correlated with $f_q$ once $f_c$ is fixed. The responses of jets in dijet and $\gamma +$ jet events are treated as independent even though this is not strictly correct. This is because the sources of jets in both samples are quarks and gluons.

Removing $f_g$, assuming that the uncertainty on $f_c$ can be ignored, and assuming that the response of jets in dijet and $\gamma +$ jet is independent allows the components of the quark and gluon response uncertainty to be added in quadrature without the need for considering correlations.
8.2 Results

The quark and gluon responses are calculated using equations 5.20 and 5.21. Components of uncertainty from the dijet quark fraction, the dijet MPF response, $\gamma +$ jet quark fraction, the $\gamma +$ jet response and the $c$ jet response have been calculated using the method described in the previous section. These components were then added in quadrature.

Figure 8.4 shows the reconstructed quark jet response in data (black), in Pythia (red) and in Herwig++ (blue) as solid circles. The MPF response of jets tagged as quark jets in Monte Carlo simulation is shown by open circles for Pythia and Herwig++. Open squares show the Monte Carlo response for jets tagged as quark jets. The Monte Carlo response, MCResponse, is the reconstructed, pile-up corrected, jet $p_T$ divided by the MC particle level jet $p_T$. Above 150 GeV the reconstructed quark response in data, Pythia and Herwig++ agree within 0.02 and also agree with the “true” response, shown by the open circles and squares, within 0.02. Below 150 GeV the reconstructed quark response in data, Pythia and Herwig++ agree within 0.04. In this region there is some disagreement between the reconstructed quark responses and the “truth” responses. Further study will be needed to understand this difference and may be a limitation on the method. The uncertainties on the calculated responses are between 0.02 and 0.05. The increase in uncertainties above 400 GeV is due to the fact that the difference in quark and gluon fractions between dijet and $\gamma +$ jet decreases in this region.

Figure 8.5 shows the gluon jet response reconstructed in data (black), in Pythia (red) and in Herwig++ (blue) as solid circles. The MPF response of jets tagged as gluon jets is shown by open circles for Pythia and Herwig++. Open squares show the Monte Carlo response for jets tagged as gluon jets. The reconstructed gluon response in data agrees with the reconstructed gluon response in Pythia within 0.01-0.05. The agreement between reconstructed gluon response in data and Herwig++ is not as good, ranging from 0.05-0.1. The reconstructed gluon responses in data, Pythia and Herwig++ are 0.03-0.1 lower than the “truth” gluon response given by the open circles and open squares in Figure 8.5. The uncertainties on the calculated responses range from 0.15-0.35. The uncertainties in the gluon jet response are larger than the quark-jet response because the gluon jet response uncertainty depends more on the dijet response uncertainty than the quark response uncertainty does. This can be seen in the equations for the gluon response uncertainty in Appendix A. The difference between the calculated gluon jet response in data and the calculated gluon jet response in simulation is likely due to the miscalibration of the reference jet energy in dijet events as well as a mismodelling of the quark and gluon fractions in dijet and $\gamma +$ jet events.
Figure 8.4: The calculated quark jet response (solid circles) for data (black), Pythia (red) and Herwig++ (blue). The open circles show the MPF response when tagging quark jets in Pythia (red) and Herwig++ (blue). The open squares show the MCRresponse in Pythia (red) and Herwig++ (Blue) when tagging quark jets.
Figure 8.5: The calculated gluon jet response (solid circles) for data (black), Pythia (red) and Herwig++ (blue). The open circles show the MPF response when tagging gluon jets in Pythia (red) and Herwig++ (blue). The open squares show the MCResponse in Pythia (red) and Herwig++ (blue) when tagging gluon jets.
Chapter 9

Conclusion

Jets are the most abundantly produced objects at the LHC. Jets which are initiated by quarks (quark jets) and jets initiated by gluons (gluon jets) have a different response. Gluon jets tend to have a lower response than quark jets because they contain a higher number of low energy particles than quark jets. To measure the response of quark jets and gluon jets separately two samples with sufficiently different quark and gluon jet fractions are required. In this thesis the samples used are dijet and $\gamma +$ jet events.

A method for measuring the response of jets in dijets was developed using the Missing transverse energy Projection Fraction (MPF) technique. The method uses a likelihood function to apply a correction to the energy of one of the jets in a dijet event so that it may be used as a reference object in the MPF technique. After applying the correction, the response of jets in dijet events was measured using 21.3 fb$^{-1}$ of $\sqrt{s} = 8$ TeV data recorded in 2012. The response predicted by Monte Carlo simulation was also determined.

The measured dijet response is consistent with the predicted response from Monte Carlo simulations. The dijet response in data and Pythia agree within 0.05 over all momentum ranges and in many cases the agreement is better than 0.02. The uncertainties on the dijet response is quite large (0.1-0.2) due to the uncertainty on the tag jet momentum. The uncertainty on the tag jet momentum comes from the resolution of the $\alpha$ closure test which ranges from 0.17 at low $p_T$ to 0.04 at high $p_T$. Further work will be needed to reduce these uncertainties. Using the measured $\gamma +$ jet and dijet responses a quark jet response and a gluon jet response were calculated.

The quark jet response reconstructed in data agrees with the reconstructed quark response from Monte Carlo simulations within 0.01-0.05. Uncertainties on the quark jet response are relatively small, between 0.03 and 0.1. At low energy, there is a discrepancy between the reconstructed quark and truth quark response of 0.02-0.1. The gluon response reconstructed in data agrees with the reconstructed gluon response in the Pythia simulation within 0.01-0.05. A larger discrepancy of 0.05-0.1 is seen between the reconstructed gluon response in data and Herwig++. The uncertainty on the gluon jet response is larger, in
the range 0.15-0.35. This large uncertainty comes from the large uncertainty on the tag jet \( p_T \) in dijet events. The gluon jet uncertainties depend more upon the dijet uncertainties than the \( \gamma + \) jet uncertainties because there is a small gluon fraction in \( \gamma + \) jet events. At high \( p_T \) the gluon fraction in dijet events also decreases, increasing the gluon jet response uncertainty at high \( p_T \).

The quark and gluon responses measured in data in this thesis will allow theorists to compare and improve the quark and gluon responses predicted by Monte Carlo models. The goal of using separate quark and gluon jet responses to reduce the jet energy scale uncertainty in samples with large gluon fraction is currently unachievable due to the large gluon jet response uncertainties. Improvements to the likelihood method for correcting the energy of a jet in a dijet event could make this possible in the future. Future work should focus on improving the resolution of the alpha closure as this will decrease the dijet response uncertainty. This in turn will decrease the gluon jet response making it feasible to use separate quark and gluon jet responses to decrease the total jet energy scale uncertainty.
Bibliography


Appendix A

Quark and Gluon Response Uncertainties

To calculate the Quark and Gluon response uncertainties, how the quark and gluon responses change as a function of their inputs must be known. If it can be assumed that the inputs are uncorrelated then the component of the uncertainty from each input may be added in quadrature. To achieve this $f_{\gamma J}^q$ and $f_{DJ}^g$ are replaced in equations 5.20 and 5.21 using the relation $f^g + f^q + f^c = 1$,

$$R^q = \frac{(1 - f_{DJ}^q) \left( R_{\gamma J} - R^c f_{\gamma J}^c \right) - \left( 1 - f_{\gamma J}^q - f_{\gamma J}^c \right) (R_{DJ} - R^c f_{DJ}^c)}{f_{\gamma J}^q (1 - f_{DJ}^q - f^c) - f_{DJ}^q \left( 1 - f_{\gamma J}^q - f_{\gamma J}^c \right)}$$  \hspace{1cm} (A.1)

and

$$R^g = \frac{f_{\gamma J}^q (R_{DJ} - R^c f_{DJ}^c) - f_{DJ}^q \left( R_{\gamma J} - R^c f_{\gamma J}^c \right)}{f_{\gamma J}^q (1 - f_{DJ}^q - f^c) - f_{DJ}^q \left( 1 - f_{\gamma J}^q - f_{\gamma J}^c \right)}.$$  \hspace{1cm} (A.2)

These equations may be written as,

$$R^q = \frac{(1 - f_{DJ}^q) \left( R_{\gamma J} - R^c f_{\gamma J}^c \right) - \left( 1 - f_{\gamma J}^q - f_{\gamma J}^c \right) (R_{DJ} - R^c f_{DJ}^c)}{f_{\gamma J}^q (1 - f_{DJ}^q - f^c) - f_{DJ}^q \left( 1 - f_{\gamma J}^q - f_{\gamma J}^c \right)}$$  \hspace{1cm} (A.3)

and

$$R^g = \frac{f_{\gamma J}^q (R_{DJ} - R^c f_{DJ}^c) - f_{DJ}^q \left( R_{\gamma J} - R^c f_{\gamma J}^c \right)}{f_{\gamma J}^q (1 - f_{DJ}^q - f^c) - f_{DJ}^q \left( 1 - f_{\gamma J}^q - f_{\gamma J}^c \right)}.$$  \hspace{1cm} (A.4)

Let

$$D = f_{\gamma J}^q (1 - f_{DJ}^q) - f_{DJ}^q \left( 1 - f_{\gamma J}^q \right).$$  \hspace{1cm} (A.5)

Assuming the uncertainty on $f^c$ is small in both the dijet and $\gamma +$ jet samples the components of the quark response uncertainty come from $f_{DJ}^q$, $f_{\gamma J}^q$, $R_{DJ}$, $R_{\gamma J}$ and $R^c$. The uncertainty components for the quark response are labelled $\Delta R^q \left( \Delta f_{DJ}^q \right)$, $\Delta R^q \left( \Delta f_{\gamma J}^q \right)$,
\[ \Delta R^q (\Delta R_{D,J}) = \Delta R^q (\Delta R_{\gamma,J}) \] and \[ \Delta R^q (\Delta R^c) \] respectively and given by,

\[
\Delta R^q (\Delta f^q_{D,J}) = \frac{\partial R^q_{\gamma,J}}{\partial f^q_{D,J}} \Delta f^q_{D,J} \\
= (-1) \left( \frac{R_{\gamma,J} - R^c f^c_{\gamma}}{D} - \left(1 - f^c_{\gamma,J} \right) \left( f^q_{D,J} \left( R_{\gamma,J} - R^c f^c_{\gamma} \right) \right) \Delta f^q_{D,J}, \right. \\
\] (A.6)

\[
\Delta R^q (\Delta f^q_{\gamma,J}) = \frac{\partial R^q_{\gamma,J}}{\partial f^q_{\gamma,J}} \Delta f^q_{\gamma,J} \\
= \left( \frac{R_{D,J} - R^c f^c_{D,J}}{D} - \left(1 - f^c_{D,J} \right) \left( f^q_{D,J} \left( R_{\gamma,J} - R^c f^c_{\gamma} \right) \right) \Delta f^q_{D,J}, \right. \\
\] (A.7)

\[
\Delta R^q (\Delta R_{D,J}) = \frac{\partial R^q_{D,J}}{\partial R_{D,J}} \Delta R_{D,J} = \frac{-f^q_{\gamma,J} \Delta R_{D,J}}{D}, \\
\] (A.8)

\[
\Delta R^q (\Delta R_{\gamma,J}) = \frac{\partial R^q_{\gamma,J}}{\partial R_{\gamma,J}} \Delta R_{\gamma,J} = \frac{f^q_{D,J} \Delta R_{\gamma,J}}{D}, \\
\] (A.9)

\[
\Delta R^q (\Delta R^c) = \frac{\partial R^c_{\gamma,J}}{\partial R_{\gamma,J}} \Delta R_{\gamma,J} = \frac{(-1) \left( f^q_{D,J} f^c_{\gamma,J} - f^q_{\gamma,J} f^c_{D,J} \right) \Delta R^c}{D}, \\
\] (A.10)

where \( f^q \) has been re-substituted into the equations where appropriate. Gluon Response uncertainties are given by,

\[
\Delta R^q (\Delta f^q_{D,J}) = (-1) \left( \frac{R_{\gamma,J} - R^c f^c_{\gamma}}{D} - \left(1 - f^c_{\gamma,J} \right) \left( f^q_{D,J} \left( R_{\gamma,J} - R^c f^c_{\gamma} \right) \right) \right. \\
\] (A.11)

\[
\Delta R^q (\Delta f^q_{\gamma,J}) = \left( \frac{R_{D,J} - R^c f^c_{D,J}}{D} - \left(1 - f^c_{D,J} \right) \left( f^q_{D,J} \left( R_{\gamma,J} - R^c f^c_{\gamma} \right) \right) \right. \\
\] (A.12)

\[
\Delta R^q (\Delta R_{D,J}) = \frac{-f^q_{\gamma,J} \Delta R_{D,J}}{D}, \] (A.13)
\[ \Delta R^q (\Delta R_{\gamma J}) = \frac{\int_D f^q_{D,J} \Delta R_{\gamma J}}{D}, \quad (A.14) \]

\[ \Delta R^q (\Delta R^c) = \frac{(-1)}{D} \left( \int_D f^q_{D,J} f^c_{\gamma J} - \int_D f^q_{D,J} f^c_{D,J} \right) \Delta R^c. \quad (A.15) \]
Appendix B

Selected Variables

This appendix lists the variables selected in each of the $p_T$ ranges.
Reco pT 25-45 GeV

**Golden**
- TrackPt500Frac
- EndLayer
- NumTowers
- EMB3OverEM
- PREBFrac
- Mass

**Intermediate**
- TrackPt500Frac
- NTrk500
- EndLayer
- MostEperX0
- NumTowers
- TILEB0Frac
- AVGTrackPt1Frac
- TrackWidth
- EMB3OverEM

**Full**
- TrackPt500Frac
- NTrk500
- EndLayer
- MostEperX0
- NumTowers
- TILEB0Frac
- AVGTrackPt1Frac
- TrackWidth
- EMB3OverEM
- PREBFrac
- EMB1Frac
- TILEB2OverTILE
- Mass
- TrackWidth

Reco pT 45-65 GeV

**Golden**
- TrackPt500Frac
- EndLayer
- NumTowers
- MostEPerX0
- TILEB1OverTILE
- EMB3OverEMB

**Intermediate**
- TrackPt500Frac
- NTrk500
- EndLayer
- TILEB0Frac
- NumTowers
- EMB1Frac
- TILEB1Frac
- MostELayerFrac
- AVGTrackPt1Frac

**Full**
- TrackPt500Frac
- NTrk500
- EndLayer
- TILEB0Frac
- EMBFrac
- NumTowers
- EMB1Frac
- TILEB1Frac
- MostELayerFrac
- MostEPerX0
- AVGTrackPt1Frac
- TILEB1OverTILE
- EMB3OverEMB
- Width

Reco pT 65-85 GeV

**Golden**
- NTrk500
- EndLayer
- EMB3OverEMB
- NumTowers
- TILEB1OverTILE
- Mass

**Intermediate**
- NTrk500
- TrackPt1Frac
- EndLayer
- TILEB0Frac
- EMB3OverEMB
- MostELayerFrac
- EMB1Frac
- TILEB1Frac
- AVGTrackPt1Frac

**Full**
- NTrk500
- TrackPt1Frac
- EndLayer
- TILEB0Frac
- EMB3OverEMB
- EMBFrac
- MostELayerFrac
- EMB2Frac
- EMB1Frac
- MostEPerX0
- TILEB1Frac
- AVGTrackPt1Frac
- NumTowers
- TILEB1OverTILE
Reco pT 85-105 GeV

Golden
NTrk500
TrackPt500Frac
EndLayer
TILEB0Frac
EMB3OverEMB
NumTowers

Intermediate
NTrk500
TrackPt500Frac
EndLayer
TILEB0Frac
MostELayerFrac
EMBOverEMB
TILEB1OverTILE
Width
MostEPerX0

Full
NTrk500
TrackPt500Frac
EndLayer
TILEB0Frac
EMB1Frac
TILEB1OverTILE
Width
MostEPerX0
TrackWidth
NumTowers
PREBFrac

Reco pT 105-125 GeV

Golden
NTrk500
EMB1Frac
TrackPt500Frac
EMB3OverEMB
NumTowers
TILEB1OverTILE

Intermediate
NTrk500
EMB1Frac
EndLayer
TrackPt500Frac
EMB3OverEMB
Width
NumTowers
TILEB1OverTILE
MostEPerX0

Full
NTrk500
EMB1Frac
TILEB1OverTILE
Width
EMB1Frac
NumTowers
TILEB1OverTILE
MostEPerX0
TrackWidth
PREBFrac

Reco pT 125-160 GeV

Golden
TILEB0Frac
EndLayer
NTrk1
TrackPt500Frac
EMB3OverEMB
NumTowers

Intermediate
TILEB0Frac
EndLayer
NTrk1
TrackPt500Frac
MostELayerFrac
EMB3OverEMB
Width
TILEB1Frac
MostEPerX0

Full
TILEB0Frac
EndLayer
NTrk1
EMB1Frac
TrackPt500Frac
MostELayerFrac
EMB3OverEMB
Width
EMB1Frac
MostEPerX0
TILEB1OverTILE
NumTowers
TrackWidth
PREBFrac
## Reco pT 160-210 GeV

### Golden
- TILEB0Frac
- NTrk1
- EndLayer
- TrackPt500Frac
- EMB3OverEMB
- NumTowers

### Intermediate
- TILEB0Frac
- NTrk1
- EndLayer
- MostELayerFrac
- TrackPt500Frac
- EMB3OverEMB
- Mass
- NumTowers
- MostEperX0

### Full
- TILEB0Frac
- NTrk1
- EndLayer
- EMBFrac
- MostELayerFrac
- TrackPt500Frac
- EMB3OverEMB
- TILEB1Frac
- Width
- NumTowers
- EMB1Frac
- MostEperX0
- TILEB1OverTILE
- PREBFrac

## Reco pT 210-260 GeV

### Golden
- TILEB0Frac
- EndLayer
- NTrk1
- TrackPt1Frac
- EMB3OverEMB
- NumTowers

### Intermediate
- TILEB0Frac
- NTrk1
- TrackPt1Frac
- EMB3OverEMB
- NumTowers
- TILEB1OverTILE
- EMB1Frac
- PREBFrac

### Full
- TILEB0Frac
- EMBFrac
- EndLayer
- MostELayerFrac
- NTrk1
- TrackPt1Frac
- EMB3OverEMB
- Width
- NumTowers
- EMB1Frac
- MostEperX0
- PREBFrac
- EMB2OverEMB

## Reco pT 260-310 GeV

### Golden
- TILEB0Frac
- EndLayer
- NTrk500
- TrackPt1Frac
- EMB3OverEMB
- NumTowers

### Intermediate
- TILEB0Frac
- EndLayer
- NTrk500
- TrackPt1Frac
- EMB3OverEMB
- NumTowers
- Mass
- PREBFrac
- TILEB1OverTILE

### Full
- TILEB0Frac
- EndLayer
- EMBFrac
- NTrk500
- TrackPt1Frac
- EMB3OverEMB
- Width
- NumTowers
- PREBFrac
- TILEB1OverTILE
- MostEperX0
- EMB1Frac
- EMB2OverEMB
- TILEB2Frac
<table>
<thead>
<tr>
<th>Reco pT 310-400 GeV</th>
<th>Intermediate</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golden</td>
<td>TILEB0Frac</td>
<td>TILEB0Frac</td>
</tr>
<tr>
<td>TILEB0Frac</td>
<td>EndLayer</td>
<td>EndLayer</td>
</tr>
<tr>
<td>EndLayer</td>
<td>EMB3OverEMB</td>
<td>EMB3OverEMB</td>
</tr>
<tr>
<td>EMB3OverEMB</td>
<td>TrackPt1Frac</td>
<td>TrackPt1Frac</td>
</tr>
<tr>
<td>TrackPt1Frac</td>
<td>NTrk1</td>
<td>NTrk1</td>
</tr>
<tr>
<td>NTrk1</td>
<td>NumTowers</td>
<td>NumTowers</td>
</tr>
<tr>
<td>NumTowers</td>
<td>Mass</td>
<td>Mass</td>
</tr>
<tr>
<td></td>
<td>TILEB1OverTILE</td>
<td>TILEB1OverTILE</td>
</tr>
<tr>
<td></td>
<td>PREBFrac</td>
<td>PREBFrac</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reco pT 400-500 GeV</th>
<th>Intermediate</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golden</td>
<td>TILEB0Frac</td>
<td>TILEB0Frac</td>
</tr>
<tr>
<td>TILEB0Frac</td>
<td>EndLayer</td>
<td>EndLayer</td>
</tr>
<tr>
<td>EndLayer</td>
<td>EMB3OverEMB</td>
<td>EMB3OverEMB</td>
</tr>
<tr>
<td>EMB3OverEMB</td>
<td>NTrk1</td>
<td>NTrk1</td>
</tr>
<tr>
<td>NTrk1</td>
<td>TrackPt500Frac</td>
<td>TrackPt500Frac</td>
</tr>
<tr>
<td>TrackPt500Frac</td>
<td>NumTowers</td>
<td>NumTowers</td>
</tr>
<tr>
<td>NumTowers</td>
<td>PREBFrac</td>
<td>PREBFrac</td>
</tr>
<tr>
<td></td>
<td>Mass</td>
<td>Mass</td>
</tr>
<tr>
<td></td>
<td>MostEperX0</td>
<td>MostEperX0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reco pT 500-600 GeV</th>
<th>Intermediate</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>TILEB0Frac</td>
<td>TILEB0Frac</td>
<td>TILEB0Frac</td>
</tr>
<tr>
<td>EndLayer</td>
<td>EndLayer</td>
<td>EndLayer</td>
</tr>
<tr>
<td>EMB3OverEMB</td>
<td>EMB3OverEMB</td>
<td>EMB3OverEMB</td>
</tr>
<tr>
<td>TrackPt1Frac</td>
<td>TrackPt1Frac</td>
<td>TrackPt1Frac</td>
</tr>
<tr>
<td>Width</td>
<td>Width</td>
<td>Width</td>
</tr>
<tr>
<td>NumTowers</td>
<td>NumTowers</td>
<td>NumTowers</td>
</tr>
<tr>
<td></td>
<td>MostEperX0</td>
<td>MostEperX0</td>
</tr>
<tr>
<td>TILEB1OverTILE</td>
<td>TILEB1OverTILE</td>
<td></td>
</tr>
<tr>
<td>EMB2OverEMB</td>
<td>EMB2OverEMB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PREBFrac</td>
<td>PREBFrac</td>
</tr>
<tr>
<td>TILEB2Frac</td>
<td>TILEB2Frac</td>
<td>TILEB2Frac</td>
</tr>
</tbody>
</table>
### Reco pT 600-800 GeV

**Golden**
- TILEB0Frac
- EndLayer
- EMB3OverEMB
- NTrk1
- TrackPt1Frac
- NumTowers

**Intermediate**
- TILEB0Frac
- EndLayer
- EMB3OverEMB
- NTrk1
- TrackPt1Frac
- NumTowers
- MostEperX0
- TILEB1OverTILE
- PREBFrac

**Full**
- TILEB0Frac
- MostELayerFrac
- EndLayer
- EMBFrac
- EMB3OverEMB
- NTrk1
- Width
- TrackPt1Frac
- NumTowers
- MostEperX0
- TILEB1OverTILE
- PREBFrac
- TILEB2Frac
- EMB2OverEMB

### Reco pT 800-1100 GeV

**Golden**
- TILEB0Frac
- EndLayer
- EMB3OverEMB
- TrackPt500Frac
- Width
- NumTowers

**Intermediate**
- TILEB0Frac
- EndLayer
- EMB3OverEMB
- TrackPt500Frac
- Width
- MostEperX0
- NumTowers
- PREBFrac
- TILEB1OverTILE

**Full**
- TILEB0Frac
- EMBFrac
- EndLayer
- EMB3OverEMB
- TrackPt500Frac
- Width
- NTrk1
- MostEperX0
- NumTowers
- TILEB2Frac
- PREBFrac
- TILEB1OverTILE
- EMB2OverEMB
- EMB1Frac

### Reco pT 1100-1500 GeV

**Golden**
- TILEB0Frac
- EndLayer
- EMB3OverEMB
- TrackPt500Frac
- NumTowers
- Width

**Intermediate**
- TILEB0Frac
- EndLayer
- EMB3OverEMB
- MostEperX0
- TrackPt500Frac
- NumTowers
- Width
- TILEB2Frac
- PREBFrac

**Full**
- EMBFrac
- TILEB0Frac
- EndLayer
- EMB3OverEMB
- MostEperX0
- TrackPt500Frac
- NumTowers
- Width
- NTrk1
- TILEB2Frac
- TILEB1OverTILE
- EMB2OverEMB
- PREBFrac
- EMB1Frac
Appendix C

Kinematic Variables

This chapter shows the distributions of the $\Delta \phi(\text{ref}, \text{jet})$ as well as the fraction of the leading jet $p_T$ carried by the third-leading jet.
Arbitrary Units

Data

Pythia

Herwig++

ptRef_105-125GeV
(leadingJet, subleadingJet)

φ
Δ
radians

Fraction of leading Jet pT

Arbitrary Units

Data

Pythia

Herwig++

ptRef_125-160GeV

Third Jet pT / lead Jet pT

Fraction of leading Jet pT
Appendix D

Variable Distributions

This appendix shows the distributions of the jet variables considered in $p_T$ ranges from 25-800 GeV.
<table>
<thead>
<tr>
<th>Layer</th>
<th>Description</th>
<th>Units</th>
<th>Data</th>
<th>Pythia</th>
<th>Herwig++</th>
<th>Reco pT 125-160 GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>MostELayer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MostELayerFrac</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MostEperX0Layer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMFrac</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NumTowers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TILEB0Frac</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TILEB1Frac</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EMB2OverEMBT otal

Data
Pythia
Herwig++
Reco pT 400-500 GeV

EMB3OverEMBT otal

Data
Pythia
Herwig++
Reco pT 400-500 GeV

EMB1Frac

Data
Pythia
Herwig++
Reco pT 400-500 GeV

EMB2Frac

Data
Pythia
Herwig++
Reco pT 400-500 GeV

EMB3Frac

Data
Pythia
Herwig++
Reco pT 400-500 GeV

Mass

Data
Pythia
Herwig++
Reco pT 400-500 GeV

MostELayer

Data
Pythia
Herwig++
Reco pT 400-500 GeV

MostELayerFrac

Data
Pythia
Herwig++
Reco pT 400-500 GeV

147
TILEB2Frac

Arbitrary Units

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Data
Pythia
Herwig++
Reco pT 800-1100 GeV

TrackWidth

Arbitrary Units

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Data
Pythia
Herwig++
Reco pT 800-1100 GeV

NTrk500

Arbitrary Units

0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16

Data
Pythia
Herwig++
Reco pT 800-1100 GeV

NTrk1

Arbitrary Units

0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16

Data
Pythia
Herwig++
Reco pT 800-1100 GeV

TILEB0OverTotal

Arbitrary Units

0
0.05
0.1
0.15
0.2
0.25
0.3

Data
Pythia
Herwig++
Reco pT 800-1100 GeV

TILEB1OverTotal

Arbitrary Units

0
0.05
0.1
0.15
0.2
0.25
0.3

Data
Pythia
Herwig++
Reco pT 800-1100 GeV

TILEB2OverTotal

Arbitrary Units

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Data
Pythia
Herwig++
Reco pT 800-1100 GeV

Width

Arbitrary Units

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Data
Pythia
Herwig++
Reco pT 800-1100 GeV

TILEB2Frac

TILEB2OverTotal

TILEB1OverTotal

TILEB2Frac

TILEB0OverTotal

TILEB1OverTotal

TILEB2OverTotal

TILEB0OverTotal

TILEB1OverTotal

TILEB2OverTotal
Appendix E

Signal and Background Jet Distributions

This appendix shows the signal and background jet distributions for all variables considered, in all $p_T$ ranges.
Reco pT 25-45 GeV

ActiveArea

0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2

Arbitrary Units

0

0.05

0.1

0.15

0.2

0.25

0.3

Reco pT 25-45 GeV

Signal

Background

Reco pT 25-45 GeV

SamplingMax

0 2 4 6 8 10 12 14 16 18 20 22 24

Arbitrary Units

0

0.1

0.2

0.3

0.4

0.5

0.6

Reco pT 25-45 GeV

Signal

Background

Reco pT 25-45 GeV

StartLayer

0 1 2 3 4 5 6 7 8

Arbitrary Units

0

0.1

0.2

0.3

0.4

0.5

0.6

Reco pT 25-45 GeV

Signal

Background

Reco pT 25-45 GeV

TILEB0Frac

0 0.1 0.2 0.3 0.4 0.5 0.6

Arbitrary Units

0

0.02

0.04

0.06

0.08

0.1

0.12

0.14

0.16

0.18

0.2

0.22

Reco pT 25-45 GeV

Signal

Background

Reco pT 25-45 GeV

TILEB1OverTTotal

0 0.2 0.4 0.6 0.8 1 1.2

Arbitrary Units

0

0.02

0.04

0.06

0.08

0.1

0.12

0.14

0.16

0.18

0.2

0.22

Reco pT 25-45 GeV

Signal

Background

Reco pT 25-45 GeV

EMB2OverEMBTot al

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Arbitrary Units

0

0.05

0.1

0.15

0.2

0.25

Reco pT 25-45 GeV

Signal

Background

Reco pT 25-45 GeV

PREBFrac

0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5

Arbitrary Units

0

0.02

0.04

0.06

0.08

0.1

0.12

0.14

0.16

0.18

0.2

0.22

Reco pT 25-45 GeV

Signal

Background

Reco pT 25-45 GeV
Reco pT 210-260 GeV

TILEB0OverTTot

Arbitrary Units

Reco pT 210-260 GeV

SamplingMax

Arbitrary Units

Reco pT 210-260 GeV

MostEperX0

Arbitrary Units

Reco pT 210-260 GeV

EMB2OverEMBTot

Arbitrary Units

Reco pT 210-260 GeV

MostEperX0Layer

Arbitrary Units

Reco pT 210-260 GeV

AVGTrackPt1Frac

Arbitrary Units

Reco pT 210-260 GeV

ActiveAreaE

Arbitrary Units

Reco pT 210-260 GeV

Signal

Background
Reco pT 500-600 GeV

NTrk500

Signal
Background

Reco pT 500-600 GeV

NumTowers

Signal
Background

Reco pT 500-600 GeV

EMB3Frac

Signal
Background

Reco pT 500-600 GeV

Mass

Signal
Background

Reco pT 500-600 GeV

TILEB1Frac

Signal
Background

Reco pT 500-600 GeV

ActiveAreaPz

Signal
Background

Reco pT 500-600 GeV

EMB2OverEMBT otal

Signal
Background

Reco pT 500-600 GeV

MostEperX0

Signal
Background

Reco pT 500-600 GeV

214
Appendix F

Dijet Response Distributions

F.1 Pythia Dijet Response Distributions

The distribution in the top left is for 25-45 GeV and the other distributions are shown in increases $p_T$ bins from left to right. The last bin is for 2000-2500 GeV.
F.2 Herwig++ Dijet Response Distributions

The distribution in the top left is for 25-45 GeV and the other distributions are shown in increases $p_T$ bins from left to right. The last bin is for 2000-2500 GeV.
Modified Poisson: 2.50 dof
mean: (-20.36 ± 0.07)%
width: (6.30 ± 0.2)%

Modified Poisson: 3.96 dof
mean: (-19.00 ± 0.04)%
width: (5.85 ± 0.04)%
F.3 Data Dijet Response Distributions

The distribution in the top left is for 105-125 GeV and the other distributions are shown in increases $p_T$ bins from left to right. The last bin is for 800-1100 GeV.
Modified Poisson: 2.67
\( \frac{dof}{2} \chi^2 \) 0.11\% ± mean: (-23.27 ± 0.12)%
± width: (8.42 ± 0.2)%

Modified Poisson: 17.19
\( \frac{dof}{2} \chi^2 \) 0.04\% ± mean: (-23.26 ± 0.05)%
± width: (7.55 ± 0.2)\%
Appendix G

Herwig++ Dijet Systematics

This chapter shows the dijet systematic uncertainties in Herwig++ simulations.

Figure G.1: Relative response difference between the nominal and varied ISR/FSR cuts in Herwig++ dijet events
Figure G.2: Relative response difference between the nominal 10% signal-70% background likelihood function and the varied 10% signal-50% background, 20% signal-50% background and 20% signal-70% background likelihood functions in Herwig++ dijet events.

Figure G.3: The relative response differences when using the nominal 14 variable log-likelihood functions and the varied 6 and 9 variable log-likelihood functions in Herwig++ events.
Appendix H

Correlations

H.1 Data Correlation Factors

This chapter shows the correlation factors between variables for data, Pythia8 and Herwig++. 
Data Correlation Factors ptRef_25-45 GeV
Data Correlation Factors ptRef_85-105 GeV
Data Correlation Factors ptRef_400-500 GeV
H.2 Pythia Correlation Factors
pythia Correlation Factors ptRef_25-45 GeV
pythia Correlation Factors ptRef_45-65 GeV
|                  | TrackWidth | Mass | TILEB2OverTotal | TILEB1OverTotal | TILEB0OverTotal | NTrk1 | NTrk500 | Width | TILEB2Frac | TILEB1Frac | TILEB0Frac | NumTowers | MostEpeX0 | MostELayerFrac | EMB3OverEM | EMB2OverEM | EMB1OverEM | AVGTrackPt500Frac | AVGTrackPt1Frac | EMB3Frac | EMB2Frac | EMB1Frac | PREBFrac | TrackPt500Frac | TrackPtFrac1 | TrackPt500Frac | EndLayer |
|------------------|------------|------|-----------------|-----------------|-----------------|-------|---------|-------|-----------|-----------|-----------|-----------|-----------|-----------------|-------------|-----------|-----------|-------------|----------|------------|-------------|-------------|-----------|
| **pythia Correlation Factors ptRef_65-85 GeV** |            |      |                 |                 |                 |       |         |       |           |           |           |           |           |                 |             |           |           |             |           |            |             |             |           |
pythia Correlation Factors ptRef_85-105 GeV
pythia Correlation Factors ptRef_105-125 GeV
pythia Correlation Factors ptRef_160-210 GeV
pythia Correlation Factors ptRef_210-260 GeV
pythia Correlation Factors ptRef_400-500 GeV
pythia Correlation Factors ptRef_800-1100 GeV
H.3 Herwig++ Correlation Factors
Herwigpp Correlation Factors ptRef_25-45 GeV
Herwigpp Correlation Factors ptRef_65-85 GeV
Herwigpp Correlation Factors ptRef_85-105 GeV
Herwigpp Correlation Factors ptRef_105-125 GeV
Herwigpp Correlation Factors ptRef_125-160 GeV
Herwigpp Correlation Factors ptRef_210-260 GeV
Herwigpp Correlation Factors ptRef_260-310 GeV
| TrackWidth | Mass | TILEB2OverTotal | TILEB1OverTotal | TILEB0OverTotal | NTrk1 | NTrk500 | Width | TILEB2Frac | TILEB1Frac | TILEB0Frac | NumTowers | MostEpeX0 | MostELayerFrac | EMB3OverEM | EMB2OverEM | EMB1OverEM | AVGTrackPt500Frac | AVGTrackPt1Frac | EMBFrac | EMB2Frac | EMB1Frac | PREBFrac | TrackPt1Frac | TrackPt500Frac | EndLayer |
|------------|------|-----------------|-----------------|-----------------|-------|---------|-------|----------|----------|-----------|-----------|-----------|-------------|-------------|-------------|-----------|-------------|-------------|-------------|-------------|-------------|---------|
|            |      |                 |                 |                 |       |         |       |          |          |           |           |           |             |             |             |            |             |             |            |             |             |         |

**Herwigpp Correlation Factors ptRef_310-400 GeV**
Herwigpp Correlation Factors ptRef_800-1100 GeV

The diagram visualizes the correlation factors between various parameters such as TrackWidth, Mass, NumTowers, TILEB0Frac, TILEB1Frac, TILEB2Frac, Width, NTrk500, NTrk1, TILEB0OverTotal, TILEB1OverTotal, TILEB2OverTotal, Mass, TrackWidth, and more. The color scale ranges from -0.8 to 1, indicating the strength and direction of the correlation.
Herwigpp Correlation Factors ptRef_1500-2000 GeV