

# **Development of a framework for Particle Flow Data-MC comparison using Run 2 data**

Christian Kirfel

Bachelorarbeit in Physik  
angefertigt im Physikalischen Institut

vorgelegt der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der  
Rheinischen Friedrich-Wilhelms-Universität  
Bonn

November 2016

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die Zitate kenntlich gemacht habe.

Bonn, .....  
Datum .....  
Unterschrift

1. Gutachter: Prof. Dr. Ian C. Brock
2. Gutachterin: ?

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theoretical and experimental basics</b>	<b>3</b>
2.1	The Standard Model of Particle Physics	3
2.2	The LHC and ATLAS	5
2.2.1	The LHC	5
2.2.2	The ATLAS Detector	6
2.2.3	The ATLAS coordinate system	6
2.2.4	Tracking detectors	7
2.2.5	Calorimeter	8
<b>3</b>	<b>Particle Flow Reconstruction</b>	<b>11</b>
3.1	The Particle Flow Algorithm	11
3.1.1	Track selection	12
3.1.2	Clusters	12
3.1.3	Matching track to cluster	12
3.1.4	Cell Subtraction	13
3.1.5	Eflow Rec performance studies	13
3.1.6	Recent updates in eflowrec	13
<b>4</b>	<b>Analysis framework</b>	<b>17</b>
4.1	Selection of good events	17
4.1.1	The Good Run List	17
4.1.2	Event cleaning	17
4.2	Trigger Tools	18
4.2.1	Trigger system	18
4.2.2	Trigger matching	18
4.3	Monte Carlo Re-weighting and scale factors	19
4.4	Object calibration and selection	19
4.4.1	Jet cleaning	19
4.4.2	Jet Calibration and Smearing	19
4.4.3	The jet vertex tagger	20
4.4.4	Muon Calibration and Selection	20
4.4.5	Electron Calibration and Selection	21
<b>5</b>	<b>Studies in Particle Flow data-MC comparison</b>	<b>23</b>
5.1	The $Z \rightarrow \mu\mu$ MC-sample	23
5.2	Event selection	23

5.3	Kinematic variables of the muons . . . . .	24
5.4	Reconstruction of the Z-Boson . . . . .	24
5.5	Kinematic variables of the recoiling jets . . . . .	25
5.6	Conclusion of the performance . . . . .	26
<b>A</b>	<b>Useful information</b>	<b>29</b>
	<b>List of Figures</b>	<b>31</b>
	<b>List of Tables</b>	<b>33</b>

# CHAPTER 1

---

## Introduction

---

The Large Hadron Collider (LHC) is an accelerator at CERN near Geneva. One of its main experiments is the general-purpose detector ATLAS. This thesis describes the performance of a reconstruction algorithm named Particle Flow.

The Particle Flow jets have recently been stabilised as one of the official jet collections to be used in the ATLAS analysis. Particle Flow combines tracker and calorimeter information and studies have demonstrated that it is a promising approach improving the angular resolution and transversal momentum resolution especially for lower transversal momentum. The aim of this thesis was to create a framework to study Particle Flow performance on 2016 13 TeV data and to do data-Monte Carlo comparison.

This thesis is structured as follows:

Chapter 2 gives a brief introduction to the standard model of particle physics, that describes the fundamental particles and their interactions. Furthermore the chapter includes a simple description of the ATLAS detector and a more detailed explanation of tracking detectors and calorimeters since they are important for Particle Flow.

The third chapter describes the Particle Flow algorithm in detail and also presents a brief overview of the Run 1 results as well as the changes that have been applied to the algorithm for Run 2.

Chapter four summarizes the analysis framework that has been developed during this thesis. It gives an explanation of all the important tools used in the framework and concludes by listing the tools that still have to be implemented or generated for Particle Flow.

Chapter 5 finally presents the results derived from data/Monte Carlo comparison for 2016 data on  $Z \rightarrow \mu\bar{\mu}$  events.



# CHAPTER 2

---

## Theoretical and experimental basics

---

### 2.1 The Standard Model of Particle Physics

The Standard Model of particle physics summarizes the current knowledge of fundamental particles and their interactions. The model applies to scales of 1 fm and below. Gravity, being the fourth fundamental force is not included as it is negligible for most phenomena at this scale.

The current view is that all matter is made out of three kinds of elementary particle being leptons quarks and mediators. There are six leptons falling into three families according to their charge, electron number, muon number and tau number.

Similar to that there are six flavors of quarks separated by strangeness (S), charm (C), beauty (B), and truth (T). As the leptons the quarks fall into three generations. For both kinds of particles the mass rises with the generations and each generation comes as a doublet. The first particle of each lepton doublet is uncharged and referred to as a neutrino while the second particle has charge  $-1$ . For each quark doublet there is an element with fractional charge  $-\frac{1}{3}$  and an element with fractional charge  $\frac{2}{3}$ . To each of these particles exists an anti particle of opposite charge.

The third kind of particle included in the standard model is the mediator. Mediators are gauge bosons the exchange of which allows the particles to interact. There are four kinds of elementary interactions of which the strong electromagnetic and weak interaction are included in the model. The fourth interaction is the gravitational force. The gauge particles for the strong interaction are the gluons carrying colour charge, the electromagnetic mediator is the photon ( $\gamma$ ) and the weak mediators are the  $W^\pm$  and  $Z$  bosons. Tables 2.1, 2.2 and 2.3 summarize the particles and their important properties.

	symbol	Charge	$L_e$	$L_\mu$	$L_\tau$
First generation{	$e$	-1	1	0	0
	$\nu_e$	0	1	0	0
Second generation{	$\mu$	-1	0	1	0
	$\nu_\mu$	0	0	1	0
Third generation{	$\tau$	-1	0	0	1
	$\nu_\tau$	0	0	0	1

Table 2.1: Lepton properties

	Symbol	Charge Q	mass [GeV]	D	U	S	C	B	T
First generation {	$d$	$-\frac{1}{3}$	4.8	-1	0	0	0	0	0
	$u$	$\frac{2}{3}$	2.3	0	1	0	0	0	0
Second generation {	$s$	$-\frac{1}{3}$	95	0	0	-1	0	0	0
	$c$	$\frac{2}{3}$	1 275	0	0	0	1	0	0
Third generation {	$b$	$-\frac{1}{3}$	4 180	0	0	0	0	-1	0
	$t$	$\frac{2}{3}$	173 210	0	0	0	0	0	1

Table 2.2: Quark properties

Interaction	Theory	Mediator	Charge	Coupling
Strong	QCD	gluons (8)	colour	1
Electromagnetic	QED	photon $\gamma$	electric charge	$10^{-1}$
Weak	GSW	$W^\pm, Z$	weak isospin	$20^{-6}$

Table 2.3: Mediator properties

Given this the standard model of particle physics has been a very successful model for a very long time and still holds for most cases. Nevertheless the model has some commonly known weaknesses and does not claim to be complete. For example the gravitational force is not included and in the standard model neutrinos are massless which would not allow the oscillations observed in neutrinos originating from the sun. For further information check [**griffith08**], [**thomson13**] and [**brock11**].

In 2012 the Higgs boson has been discovered at the Large Hadron Collider. It is a spin-0 scalar particle with a mass of  $m_H = 125 \text{ GeV}$  and it represents the mechanism which gives all particles their mass.

## 2.2 The LHC and ATLAS

The analysis for this thesis has been performed in the ATLAS collaboration. The ATLAS-Detector is one of the four main experiments at the LHC at Cern. This section provides a brief overview of the LHC and ATLAS detector focusing on the aspects directly relevant for Particle Flow analysis.

In addition to a brief description of the ATLAS detector a more detailed explanation of tracker and calorimeter is given since these components are directly relevant for the explanation of Particle Flow.

### 2.2.1 The LHC

The Large Hadron Collider ("LHC") located at the facilities of the European Organization of Nuclear Research ("CERN") close to Geneva was built to extend the frontiers of modern particle physics by delivering high luminosities and reaching unprecedented high energies. The hadronic collider has a circumference of about 27 km and is located on average 100 m underground.

The LHC is designed to collide bunches of up to  $10^{11}$  protons at a luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . The beams collide at four points where the four main experiments of the LHC are located. Two of these are special-purpose detectors, namely LHCb and ALICE while the other two, ATLAS and CMS, are general-purpose detectors. The analysis in this thesis was performed on ATLAS data. Figure 2.1 shows the LHC, the four detectors and its general location.

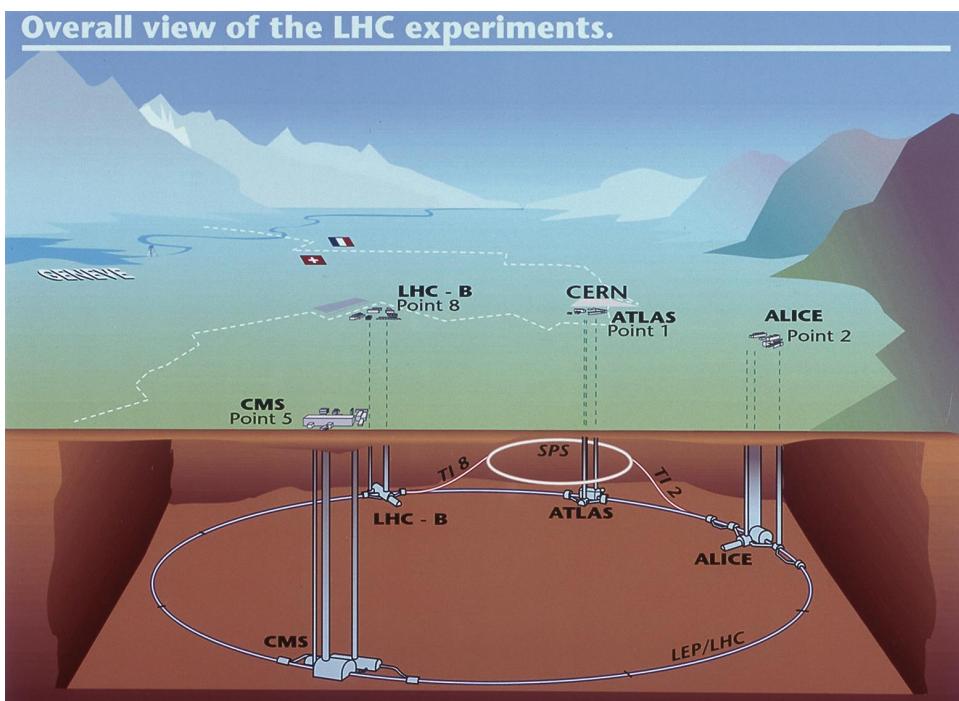


Figure 2.1: Sketch of the LHC ring, the position of the experiments and the surrounding countryside. The four big LHC experiments are indicated(ATLAS, CMS, LHC-B and ALICE)along with their injection lines(Point 1, 2, 4, 8)[[atlasfigures](#)]

### 2.2.2 The ATLAS Detector

The ATLAS detector was developed to study the physics processes in a broad energy range available at the LHC. This enables the observation of highly massive particles that lower energy accelerators were not able to create and that would bring new physics theory beyond the standard model of particle physics. It was designed to cover the maximum number of final stages being a so called general purpose-detector. Figure 2.2 is a sketch of the ATLAS detector together with a rough scale in size not only by the given dimensions on the top and left side but also by including two average sized humans close to the left muon chambers. In the following explanations of its components are given from the inside to the outside.

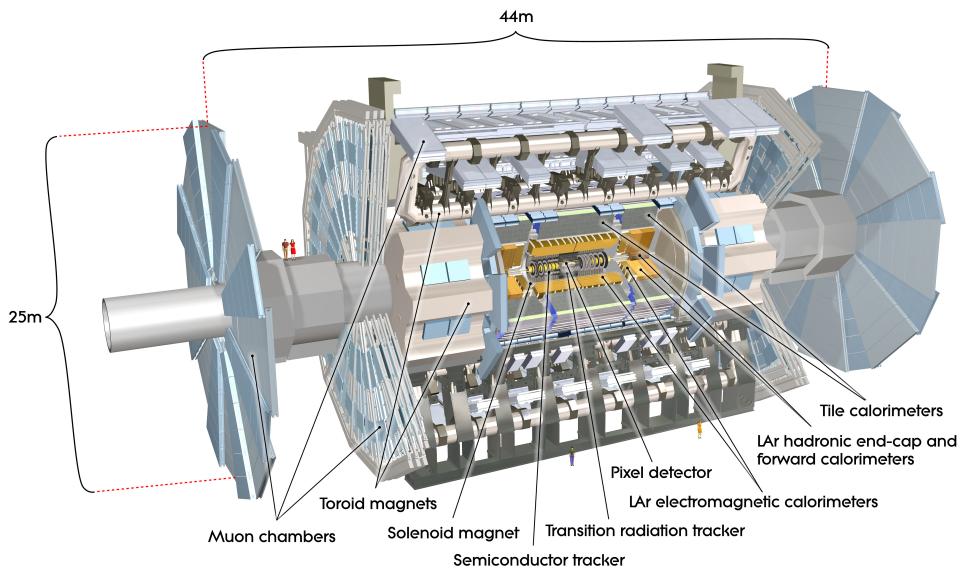


Figure 2.2: Sketch of the ATLAS detector [[atlasfigures](#)]

Figure 2.3 shows the detector's components in a simplified way and allows to understand the importance of the order of the detector's parts. The innermost part of the detector is a tracking detector surrounded by a solenoid that creates a magnetic field to bend the charged particles trajectory and measure their charge and momentum. The following part of the detector is the calorimetry system. It consists of an inner electromagnetic calorimeter and an outer hadronic calorimeter. The outermost part is a muon spectrometer because most of the particles that cross the calorimeters undetected or do not deposit their complete energy are muons.

The detector system therefore allows to measure charge, momentum and energy of most particles.

### 2.2.3 The ATLAS coordinate system

The ATLAS coordinate system is defined by the beam direction with the  $z$ -axis pointing along the LHC's beam pipe. The corresponding transverse plane is defined by the  $x$ -axis pointing towards the ring's centre. The  $y$ -axis points upwards. The origin of the system is defined by the nominal point of interaction. The polar angle,  $\theta$ , is the angle between the  $z$ -axis and the  $x$ - $y$ -plane and the azimuthal angle,  $\phi$  is the angle between the  $x$ - and the  $y$ -axis.

The coordinates used in this thesis are usually the azimuthal angle,  $\phi$ , the pseudo-rapidity,  $\eta$ , and the transverse momentum,  $p_T$ . The pseudo-rapidity replaces the polar angle and is defined as

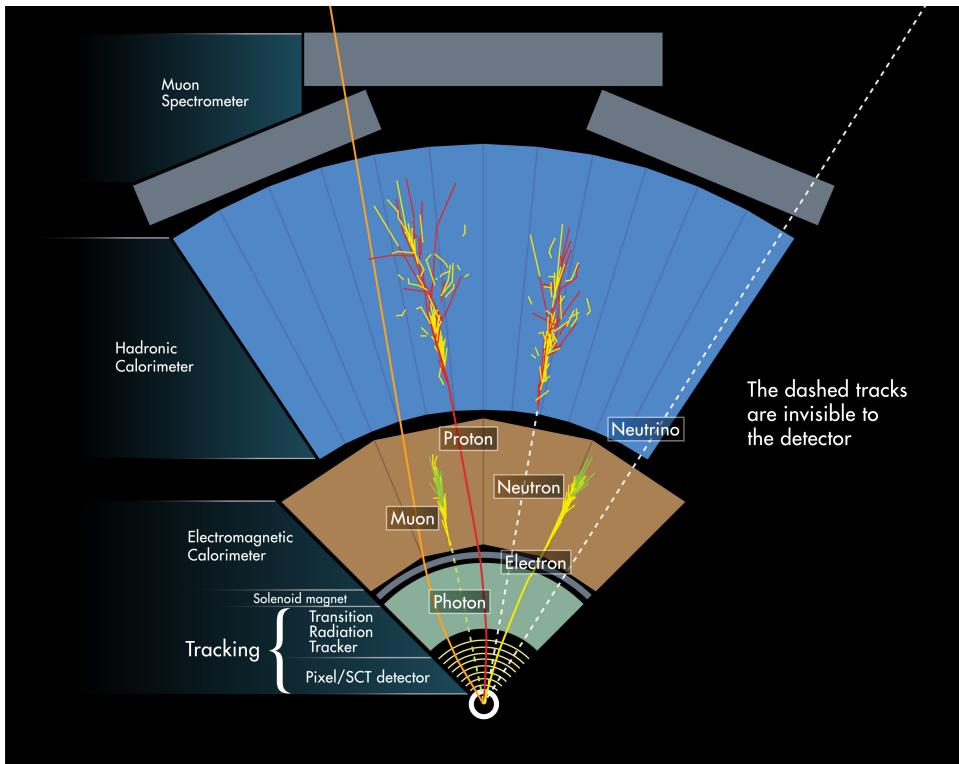


Figure 2.3: Scheme of the ATLAS-detector [atlasfigures]

$$\eta = \frac{1}{2} \ln \left[ \tan \left( \frac{\theta}{2} \right) \right]. \quad (2.1)$$

The transverse momentum is defined by

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (2.2)$$

where  $p_x$  and  $p_y$  are the momenta along the corresponding axes.

The angular variables are defined within

$$\eta \in [-\infty, \infty], \phi \in [-\pi, \pi]. \quad (2.3)$$

#### 2.2.4 Tracking detectors

To measure momentum, trajectory and charge of charged particles usually tracking detectors are used.

There are two main categories of tracking detectors following the same general principle, gaseous detectors and semiconductor detectors. If ionizing radiation passes any given medium it will create electron-hole pairs. These charge carriers can then be collected by an electric field. Depending on the detector the signal caused by the charge carriers can be related to the coordinate of ionization in space and time.

- Gaseous detectors: Gaseous detectors are based on the greater mobility of ions and electrons in the gas. The basis of the detector is usually a chamber filled with a proper gas. The gas is filled with an area of wires to which a strong electric field is applied. If gas atoms get ionized the charge

carriers (electrons and ions) will drift to the wires and create a detectable signal. The wire gives a rough estimation of space which is normally improved by calculating the exact ionization location from the drift time.

- Semiconductor detectors: Semiconductor detectors are as their name implies based on crystalline semiconductor material for example silicon and germanium. Their working principle is quite similar to that of gaseous detectors but the gas is exchanged by solid semiconducting material. In semiconducting material ionizing radiation will create electron-hole pairs instead of electron ion pairs that then can travel in a strong electric field to be detected. The big advantage of semiconductor detectors over gaseous detectors is that the energy required to create an electron-hole pair is about 10 times smaller than the energy needed to ionize gas atoms. These detectors are commonly structured into wafers or pixels that allow a determination of the space.

Usually a magnetic field surrounds tracking detectors to bend the track and that way be able to compute the particles momentum and charge based on the curvature.

In the ATLAS detector both the inner detector and the muon spectrometer are tracking detectors.

### **Inner Detector**

The innermost part of the ATLAS detector is called the Inner Detector, which consists of three sub-components, the Pixel detector (Pixel), the Semi-Conductor Tracker (SCT) and a Transition Radiation Tracker (TRT). Each of these sub-detectors is divided into the so called barrel part and two end-caps. The Inner Detector covers a region of  $|\eta| < 2.5$  which also limits the region in which Particle Flow can be used as its highest efficiency.

### **Muon spectrometer**

The second tracking detector of ATLAS is the muon spectrometer which is the outermost part of the detector. The task of the spectrometer is to detect charged particles transversing the calorimeter without being stopped or deplying their complete energy, and to do both trigger and tracking to measure their momentum. Due to these two tasks the spectrometer is divided into two parts. The first part is the trigger chamber covering a range of  $|\eta| < 2.4$ , followed by the high-precision chamber with a range of  $|\eta| < 2.7$ . The main detector's support feet cause a further gap at about  $\phi = 300^\circ$  and  $\phi = 270^\circ$ .

The high-precision detector uses monitored drift tubes (MDTs) with the exception of the innermost part of the innermost end-cap disk which utilizes Cathode Strip Chambers (CSCs). The trigger chamber uses resistive-plate chambers (RPCs) for the barrel parts and Thin-Gap Chambers (TGCs) for the end-caps. The momentum calculation is then performed by the field of the toroid magnet.

### **2.2.5 Calorimeter**

In particle physics a calorimeter is a device to measure first and foremost the total energy of a particle. Most of the time additionally some positional information is taken. The idea is that most particles loose all their momentum while crossing the calorimeter. Measuring the energy deposited this way gives a value for the particle's energy. Usually a particle deposits its energy by initiating a particle shower, the energy of which is then collected and measured. Calorimeters are distinguished by the main interaction of the particles one aims to detect.

### Electromagnetic Calorimeter

Electromagnetic calorimeters are designed to detect charged particles that primarily via the electromagnetic interaction and measure their total energy. Usually these particles are electrons and photons. There are various methods to construct these detectors. An example would be the usage of inorganic scintillators. These scintillators should be optically transparent and have a short radiation length to contain the shower in a compact region. The detection can then be followed by photon detectors with photo-multipliers which measure the emitted light being proportional to the detected particle's energy.

The electromagnetic calorimeter at ATLAS is a high-resolution and high-granularity liquid-argon sampling calorimeter with lead as absorber material. The calorimeter consists of two half-barrels which are only separated by a small gap at the interaction point. The endcaps at each side are segmented into two coaxial wheels to cover different polar angles.

### Hadronic calorimeters

Hadronic calorimeters are used to obtain the energies of hadronic particles. Due to the relatively large distance between interactions these calorimeters occupy a significantly large volume in the detector.

A common technique to construct these calorimeters is a sandwich-like structure of alternating layers of high density absorber material and active material. The absorbers are used to develop the particle showers which then hit the active material and deposit their energy there. The determination of the particle belonging to the deposited energy relies on tracker information as is sketched in figure 2.3. Energy in the hadronic calorimeter without a track implies a neutral hadron, for example a neutron. A single track paired with a energy deposition means that the particle was a charged hadron like a proton and if many tracks belong to a deposition a jet has been the most likely origin of the energy deposition.

The hadronic calorimeter system at ATLAS is divided into three calorimeter components. The first one is the scintillator-tile calorimeter covering a region of  $|\eta| < 1.7$ . The other two are the end-cap calorimeters which use liquid argon (LAr) and cover the region of  $1.5 < |\eta| < 3.2$ . The tile calorimeter itself is divided into a central barrel and two extended barrels (compare 2.2).

For more information about the ATLAS detector see the ATLAS design report [[atlastdr](#)].



# CHAPTER 3

---

## Particle Flow Reconstruction

---

One of the goals of this thesis is to build a framework to study the performance of Particle Flow jet performance in Run 2. Therefore this chapter will give an overview of the results that the new algorithm has brought for Run 1.

This chapter starts with an introduction to the Particle Flow algorithm based on the Particle Flow paper [[pflow16](#)]. After the description of the steps a brief overview of the results the algorithm brought for Run 1 are presented. The last section summarizes the updates to the algorithm for Run 2.

### 3.1 The Particle Flow Algorithm

Recently only either the Calorimeter or the tracker information was used to reconstruct jets in ATLAS. The Particle Flow algorithm now combines tracker and calorimeter information to achieve better resolution especially at lower energies. The main advantages of including the tracker information into reconstruction are as follows:

- The momentum resolution measured by the tracking detector is superior to the calorimeter for low energy charged particles.
- The tracking detector is able to reconstruct soft particles, which would not pass the noise threshold of the calorimeter and therefore not being reconstructed in the calorimeter at all.
- The ATLAS tracking detector has a superior angular resolution for single charged particles.
- Low  $p_T$  charged particles may be swept out of the cone before reaching the calorimeter by the magnetic field. The tracker information allows to cluster these particles into the jet.
- The vertex determination possible due to the use of tracks reduces the pileup-contribution considerably.

Figure 3.1 sketches the important steps of the Particle Flow algorithm. The algorithm uses clusters and tracks as input information. The first step is to match a track spatially to a cluster. After a pair has been found the algorithm checks whether the particle's momentum matches the expected energy deposited in the cluster within the expected deviation. Then:

- If the energy matches the subtraction algorithm starts deciding which calorimeter cells belong to the given jet.

- If the energy deposited in the cluster is too low the algorithm includes all the clusters in a given area and then starts the subtraction.

After the subtraction the algorithm provides not only the modified clusters with the identified remnants but also the tracks and the original unchanged clusters for further analysis.

The following sections explain each step of the algorithm in more detail.

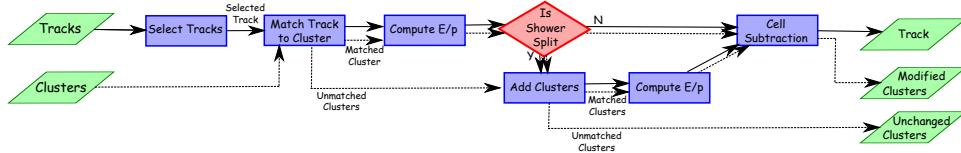


Figure 3.1: Flowchart of the steps of the Particle Flow algorithm [pf16]

### 3.1.1 Track selection

The tracks are selected if they pass certain cuts which are applied in order to minimize the amount of fake tracks. The requirements are at least 9 hits in the PIXEL plus SCT and no missing hits in the PIXEL at all. The tracks have to be in a pseudo-rapidity region of  $|\eta| < 2.5$  and  $0.5 \text{ GeV} < p_T < 40 \text{ GeV}$ .

### 3.1.2 Clusters

The calorimeter input information of the Particle Flow algorithm comes in the form of topological clusters. The construction of these clusters is briefly described here.

The idea of creating topological clusters is to group neighbouring cells that exceed the expected noise by a significant amount into collections. Each cluster is being constructed around a so called seed cell. A seed cell is a cell for which the deposited energy exceeds the expected noise by four times the standard deviation. If a seed is found all the neighbouring cells which exceed the noise by at least two times the standard deviation are added to the cluster. Finally all the cells neighbouring these cells are also added to the final cluster.

The final number of cells in a cluster is therefore not static. For more information see the report "Calorimeter Clustering Algorithms: Description and Performance" [cluster08].

### 3.1.3 Matching track to cluster

The algorithm tries to match every selected track to one single best-match calorimeter cluster. Therefore the distances in  $\Delta\phi$  and  $\Delta\eta$  from the track extrapolated to the second layer of the EM calorimeter and the topo-clusters have to be calculated. After that the topo-clusters get ranked based on the metric:

$$\Delta R' = \sqrt{\left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta\eta}{\sigma_\eta}\right)^2} \quad (3.1)$$

Where  $\sigma_\eta$  and  $\sigma_\phi$  refer to the angular topo-cluster width, computed from the standard deviation of the displacements of the topo clusters and  $\Delta\phi$  and  $\Delta\eta$  are calculated as follows:

$$\Delta\phi = (\phi_{track} - \phi_{cluster}) \Delta\eta = (\eta_{track} - \eta_{cluster}) \quad (3.2)$$

If the energy in this cluster is greater than or equal to the energy expected from the track's  $p_T$  the algorithm goes to cell subtraction. If the energy in the cluster is smaller than the expected energy all clusters in a cone of  $\Delta R < 0.2$  are matched to the track. In that case  $R$  is calculated by the metric:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \quad (3.3)$$

If the energy of all the matched clusters still does not match the expected energy the matching failed and no subtraction takes place.

### 3.1.4 Cell Subtraction

The last step in the Particle Flow algorithm is the cell-wise subtraction of energy deposits to remove noise remnants and determine which energy depositions belong to the matched track or to other neutral objects. If the energy deposited in the cluster or the set of clusters is lower than the expected energy, the clusters are simply removed as mentioned before. Otherwise, a cell by cell subtraction is performed.

The first step of the cell subtraction is to generate a shower shape from the extrapolated track. Around the extrapolated track direction rings in the  $\eta, \phi$  plane are generated just wide enough to independently contain at least one cell from the extrapolated position. Furthermore the rings are restricted to one layer and of the same radial size for each layer. After the generation of rings in each layer the average energy density in each ring is computed and the rings are ranked by energy density in descending order. The layer is not used in any way for this ranking. The subtraction then starts from the ring with highest energy density and proceeds successively to rings of lower order until the next ring's energy exceeds the remaining expected energy. If the ring's energy exceeds the energy still to be subtracted the energy in each cell is scaled down by the fraction needed to reach the expected energy before the process halts. The removed cells are kept as the corresponding track information as a charged Particle Flow object (cPFO). The cells still remaining are removed as remnants if they are consistent within the standard deviation of  $\sigma(E/p)$ . If the remaining energy is larger than  $\sigma(E/p)$  the cells are kept as a neutral Particle Flow object (nPFO). An example of the process is sketched in figure 3.2.

### 3.1.5 Eflow Rec performance studies

The results extracted from the Particle Flow paper [pflow16] show the impact of the algorithm on the angular resolution and on the rejection of pileup jets. This section briefly summarizes the improvements in the Particle Flow jet performance compared with other jet collections based on calorimeter information (LC). Figure 3.3(c) and 3.3(d) show the improvements in angular resolution while figure 3.3(b) displays the increased rejection of fake jets for the new algorithm. LC+JES jets are the jets using the old algorithm and JVJ in figure 3.3(b) refers to the Jet Vertex Fraction representing the amount of energy in the jet originating from the original vertex.

The plots clearly demonstrate that the Particle Flow algorithm does improve the angular resolution in low  $p_T$  regions while having no drawback for higher  $p_T$  regions. The pileup contribution is also mediated massively even in comparison to the usage of a cut on the JVT. The region of effect is restricted to  $|\eta| < 2.5$  because only this region of pseudorapidity is covered by the Inner Detector. Only the momentum resolution shown in figure 3.3(a) worsens using the old reconstruction for high  $p_T$  regions.

### 3.1.6 Recent updates in eflowrec

The description of the Particle Flow algorithm given in this thesis is based on the algorithm implemented in Run 1.

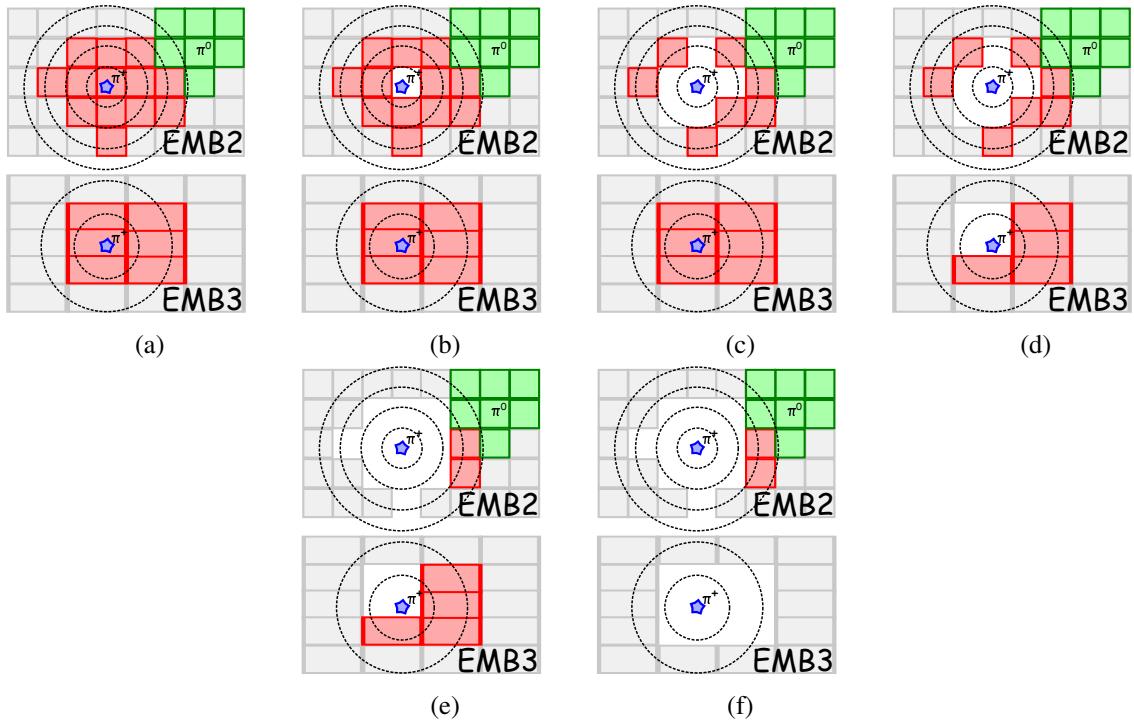
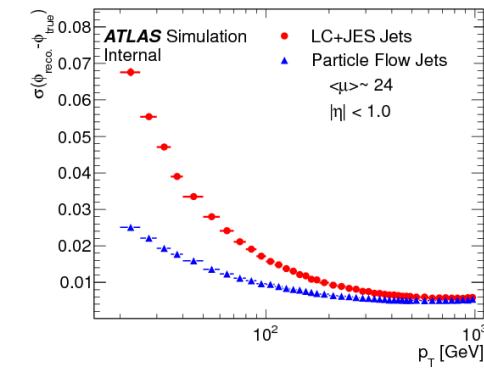
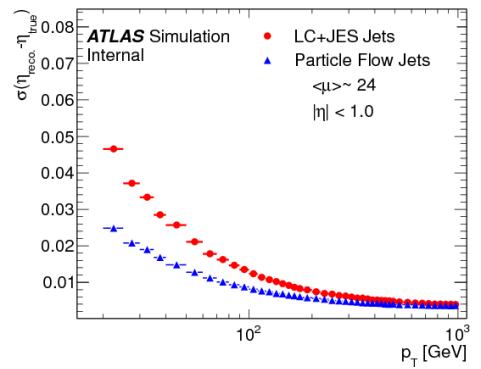
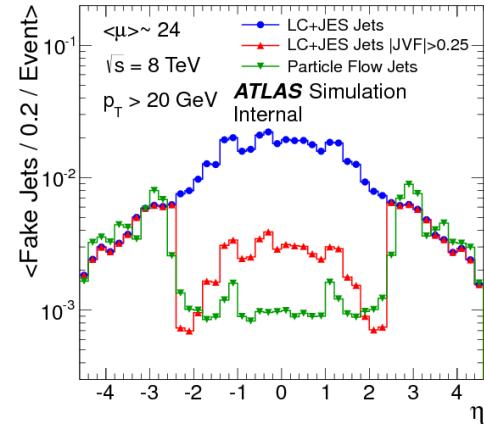
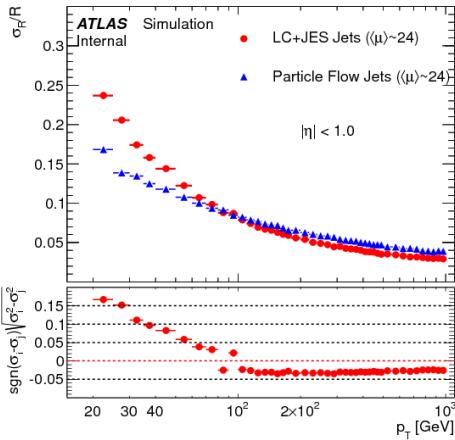


Figure 3.2: Example of cell subtraction. In red the energy deposited by the  $\pi^+$  of interest are shown and in green a cluster from a  $\pi^0$  is shown. The algorithm successfully determines the cells belonging to the  $\pi^+$  and removes them while leaving the green cells as remnants. Only in subplot 3.2(f) part of the green cells is removed while part of the red cells remain because both fall into the same subtraction ring. [pflow16]

Recently some changes have been included and they are briefly mentioned in this section:

- The track selection has been updated to the tight track criteria of the ATLAS tracking group.
- In dense environments where the association of energy deposited in the calorimeter to the track can not be done properly the subtraction is not applied.





# CHAPTER 4

---

## Analysis framework

---

In order to study the Particle Flow jet performance and the data-MC comparison a large amount of ingredients is needed to be included and checked to be correctly working. This chapter goes over the important tools included in the analysis framework. The impact of every given tool has been studied and the status of the framework as well as possible problems in the current implementation are also included. The chapter starts with the event selection. The second section describes the trigger system and its corresponding tools and finally the calibration tools for specific objects are introduced.

The final section describes the tools that still have to be implemented to achieve a better matching of data and Monte Carlo.

### 4.1 Selection of good events

The event selection is applied to mostly remove bad or corrupted data. The tools used are summarized below.

#### 4.1.1 The Good Run List

Before any analysis or calibration takes place the good run list has to be applied. The Good Run List is only needed for data and not for Monte Carlo.

For data to be suitable for analysis one has to make sure that it fits certain requirements on the detectors working state. The Good Run List allows to exclude data-taking periods in which the detector showed a poor data quality. Reasons for this may be maintenance on sub detectors, magnets off or ramping or an unstable beam of the LHC. The Good Run List includes all the good data taking periods. For the data analysis in this thesis the recommended GRL for 2016 data was used.

Figure 4.1 shows the number of events removed due to several cuts in the framework. The cuts due to the GRL are shown in the second bin. In this case no or at least very few events are removed due to the GRL. Depending on the data used the amount of events removed can be way higher.

#### 4.1.2 Event cleaning

Additionally to the cuts provided by the GRL some further events have to be excluded. Noise bursts and corrupted data in general have to be removed in the LAr, the SCT and the Tile. Furthermore due to production errors some events are duplicate and have to be also removed.

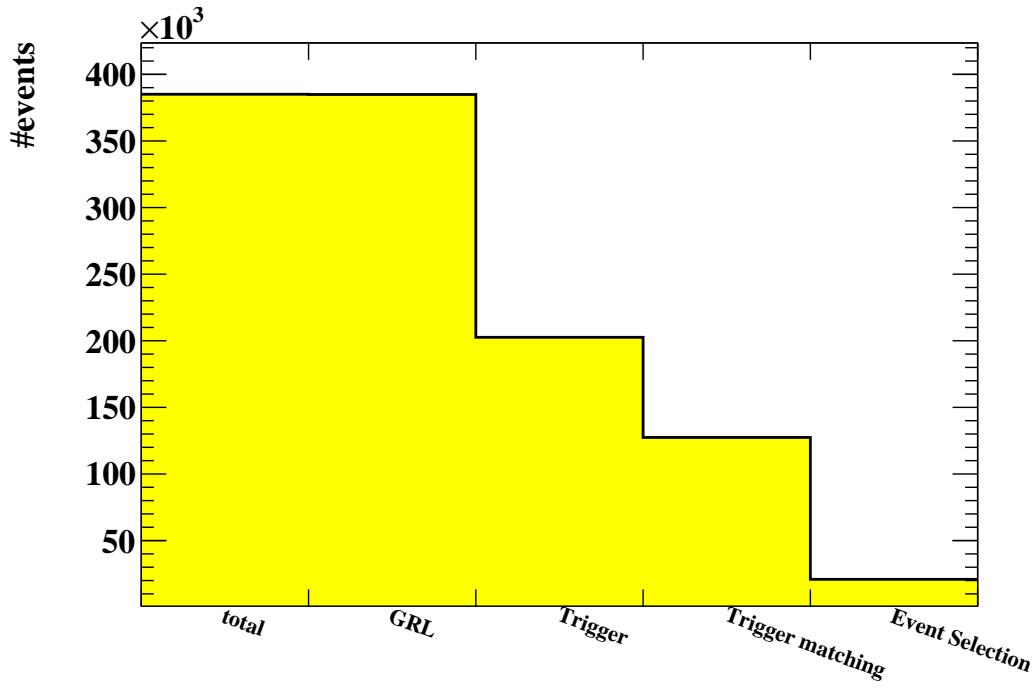


Figure 4.1: Fraction of events removed due to the GRL, trigger, trigger matching and event selection

## 4.2 Trigger Tools

The trigger system makes sure that only the events that are interesting for physics analysis are kept. A further important collection of tools has to make sure that the trigger is fired, correctly used and also that the particle that triggered is actually one of those used in further analysis.

### 4.2.1 Trigger system

A trigger basically is a first selection for an event meaning that an event is required to surpass certain demands to be used in analysis. These demands are embodied by so called trigger chains that can be used as input for a trigger tool in analysis which on that base can select or refuse events. For the analysis in this thesis the recommended single lepton triggers for 2016 data were used. The chains were HLT\_mu26\_ivarmedium, HLT\_mu50.

### 4.2.2 Trigger matching

Usually the trigger is checked before the event is further cleaned and calibrated. Therefore it can happen that the particles that passed the trigger later get removed in the analysis. The trigger matching makes sure that the particles that passed the triggers are still left in the final analysis and if not the event can still be removed.

## 4.3 Monte Carlo Re-weighting and scale factors

The Monte Carlo is produced before data is taken therefore the simulation may vary from the data for several reasons. For example the pileup distribution used to generate the MC may not match the one in data. Pileup can occur in particle collision experiments when several events overlap. There is space and time pileup. In the case of in time pileup two events happen at the same time and in space pileup refers to secondary vertices. In both cases it is important to distinguish objects from the event of interest and those originating from pileup.

To compensate these differences a sum of weights is applied to Monte Carlo as well as to data.

The most general weights are the Monte Carlo event weight and the pileup re-weighting. Nevertheless there are numerous scale factors to be applied to all kind of objects. Jets, electrons, muons, tauons and also photons all require scale factors to match the data.

All these scale factors have to be multiplied and added as a weight to every event in both data and simulation to optimize the agreement between data and Monte Carlo as far as possible.

The current framework only includes the Monte Carlo weight and the pileup re-weighting. The other weights are going to be included in future updates of the developing. Anyhow the changes expected from the missing scale factors are expected to be rather small and they do not stop the results from this work from being significant.

## 4.4 Object calibration and selection

The last step in the framework is the selection and calibration of the objects per event. This section summarizes the tools needed not only for jets but also for muons and electrons, which are the objects present in our event topology.

### 4.4.1 Jet cleaning

The jet cleaning is applied to identify bad jets. Bad jets are jets not associated to real energy deposits in the calorimeters. They can occur due to a broad range of reasons: hardware problems, LHC beam condition or even cosmic-ray showers.

The Jet Cleaning is used on both data and Monte Carlo to make sure that Monte Carlo events that would be removed in data also are not included in the simulation. Bad jets are excluded on the base of some of their properties: negative energy, charged fraction and energy deposited in specific calorimeter layers. The set of criteria is rather big and more information can be obtained in the twiki for example at [\[jetcleaning\]](#).

Jet cleaning variables are not yet available for Particle Flow jets and the cleaning has to be performed on topo-jets at the current state of the framework. If a bad topo-jet is found the whole event is removed from the analysis. For this framework a "loose" [\[jetcleaning\]](#) selection has been chosen since it is generally recommended.

### 4.4.2 Jet Calibration and Smearing

After making sure a jet is "good" and therefore has passed the cleaning it must still be calibrated and in case of MC smeared to data. The calibration scales the energy of jets for a certain reconstruction algorithm. The Smearing smears the MC resolution to be matching the actual data resolution.

Figure 4.2 shows the impact of these tools on the energy and the momentum of jets in a Monte Carlo sample.

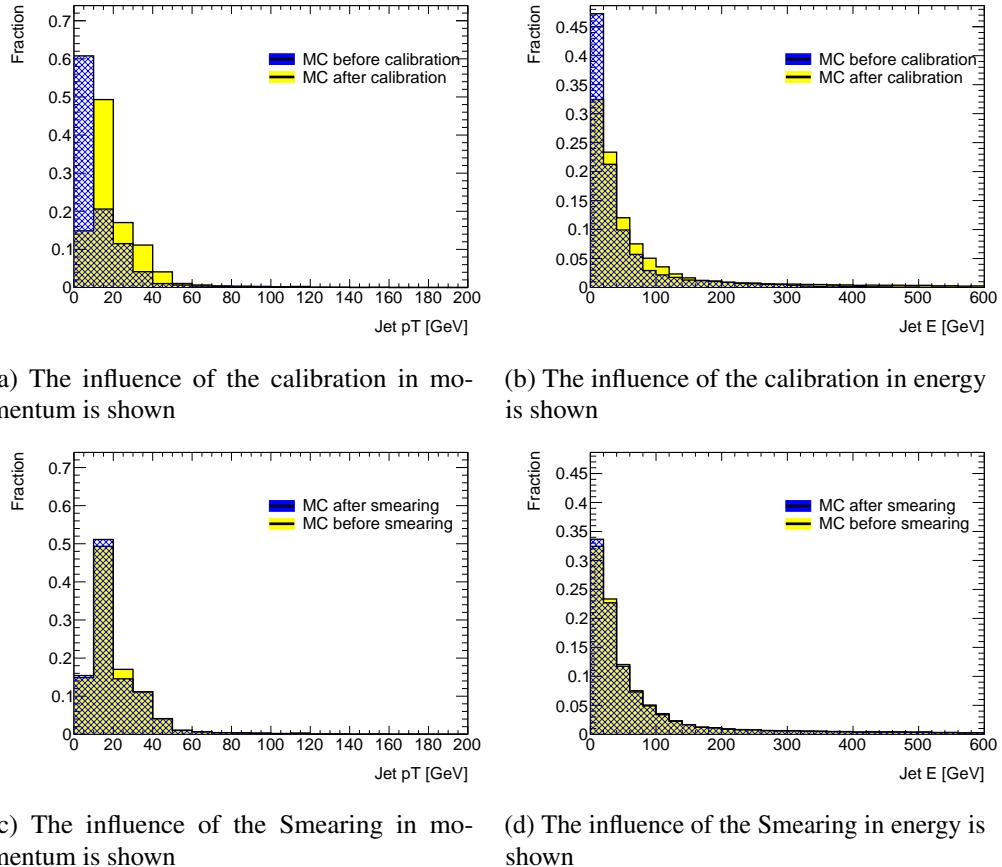


Figure 4.2: The figure demonstrates the results of calibration and smearing for Particle Flow jets. Figures (a) and (b) show the jet momentum and energy before and after scaling and figures (c) and (d) show the changes in the same variables for the smearing.

#### 4.4.3 The jet vertex tagger

Due to the high luminosity at the LHC pileup is a big challenge in ATLAS analysis. Therefore the rejection of pileup is an important part of analysis. One way to reject pileup is to calculate the jet-vertex fraction of each jet which is the fraction of momentum in the jet originating from the primary vertex. If one sets a minimum on this fraction pileup can be suppressed because jets that do not pass the criteria on their JVF are highly likely to be originating from pileup vertices. The Jet Vertex Tagger relates each jet to a vertex.

#### 4.4.4 Muon Calibration and Selection

Muons are reconstructed using both the muon spectrometers and the inner detector. The information from both detector systems is combined to a single track.

Analogue to jets the muons also have to be calibrated properly. This section introduces all the important tools for muon calibration and gives a brief summary of the effects of the cuts and calibrations. The first task of the muon tools is to determine whether a muon originated from the original vertex or has its origin in background noise( cosmic muons) or in some kind of secondary interaction.

The muons are requested to have  $p_T > 25 \text{ GeV}$  and a pseudorapidity of  $|\eta| < 2.5$ . To reject the cosmic

muon background further the muons are not allowed to have a longitudinal impact parameter to the primary vertex higher than 3 mm.

The last requirement is added due to muons originating from heavy flavour quark decays. To remove these muons an isolation criteria is implemented, namely the isolation tool. The tool is responsible to make sure that the sum of transversal momentum of the tracks around the muon candidate divided by the muons momentum is smaller than 0.05. 4.1. The efficiency of this tool has been checked with  $Z \rightarrow \mu^+ \mu^-$  events providing a 97 % in  $\mu$  detection.

$$\frac{\sum_{\Delta R} pT_{track}}{pT_\mu} < 0.05 \quad (4.1)$$

The second part of the muon tools makes sure that the muon properties are correctly calibrated and smeared to make Monte Carlo and data match properly and to eliminate known weaknesses in the detector structure. Figure 4.3 shows the muons transversal momentum before and after calibration in Monte Carlo. The changes are very minor.

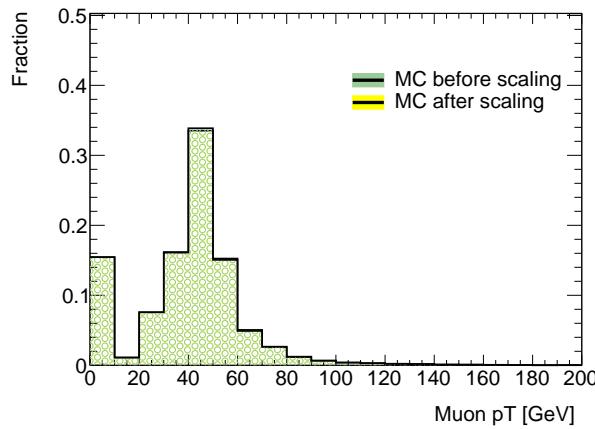


Figure 4.3: Calibration and smearing of the muon momentum before any further cuts are applied. The changes are very minimal.

#### 4.4.5 Electron Calibration and Selection

Electrons are detected by leaving a track in the inner detector and depositing lose to all of their energy in the electromagnetic calorimeter. The reconstruction algorithm expects a calorimeter cluster with a deposited energy  $E_T$  exceeding 2.5 GeV. This cluster has then to have a matching track from the primary hard scatter vertex. The  $\eta$  requirements are  $|\eta|_{cluster} < 2.47$  with an exclusion of  $1.37 < |\eta| < 1.52$  (calorimeter barrel-endcap transition region).

The electron tools are very similar to the muon tools. There is a first group of tools and criteria to determine which electrons actually originate from the primary vertex and distinguishes those electrons from those originated in background and pileup events. The second group of tools calibrates electrons properties.

Electrons are same as muons required to be isolated. The Isolation is based on a  $\Delta R < 0.2$  cone in the deposited energy and a  $\Delta R < 0.3$  cone around the track.

The electron calibration and smearing is exemplary displayed in figure 4.4. The impact of the scaling is significantly higher than for muons.

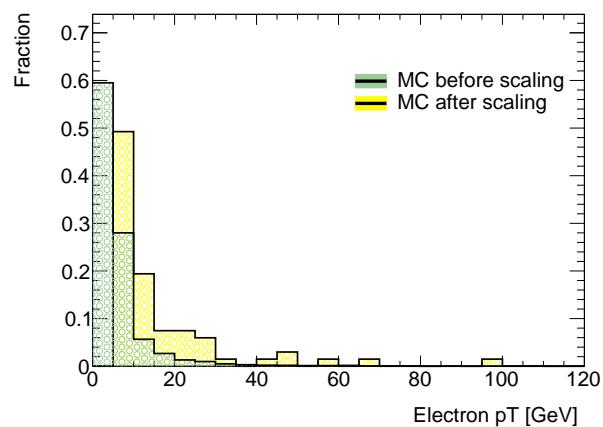


Figure 4.4: Calibration and smearing of the electron momentum.

# CHAPTER 5

---

## Studies in Particle Flow data-MC comparison

---

To study the data-MC performance of Particle Flow jets with Run 2 the  $Z \rightarrow \mu\mu$  MC-sample has been used. The chapter explains the MC sample used and the event selection before showing a summary of performance plots.

### 5.1 The $Z \rightarrow \mu\mu$ MC-sample

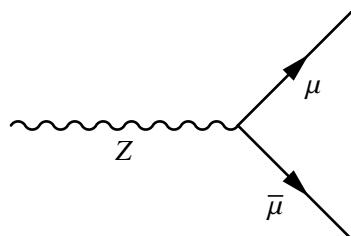


Figure 5.1: Decay of a Z-Boson to two muons

For this thesis  $Z \rightarrow \mu\mu$  events were used. The decay channel of a Z boson into a muon and an antimuon has a branching ratio of  $3.366 \pm 0.007$  [pdg]. The event was chosen for this analysis because the Z boson is very easy to trigger on and it allows a very clear event selection. Furthermore the event has exactly one recoiling jet that analysis can be performed on.

### 5.2 Event selection

An event selection is applied to enrich the sample with  $Z \rightarrow \mu\mu$  events. The cuts are defined to exploit the  $Z \rightarrow \mu\mu$  topology. The cuts applied were:

- No good electrons, where good means that the particle passed all selection filters itself.
- Exactly two good muons with opposite charge.
- $p_{T,\mu} > 25 \text{ GeV}$
- The muons are required to be in a central region of pseudo-rapidity,  $|\eta| < 2.4$ .

Figure one shows the number of events selected.

### 5.3 Kinematic variables of the muons

After selecting proper  $Z \rightarrow \mu\mu$  events the agreement between data and MC has been studied for the selected objects. This section summarizes the kinematic variables of the muons.

Figure 5.2 shows the kinematic variables of the muons. For  $pT < 100$  GeV and  $e < 300$  GeV the agreement between data and Monte Carlo for momentum and energy is very good but it is worse at a higher  $p_t$ . This might have its reason in statistics. For  $\eta$  and  $\phi$  the agreement is good.

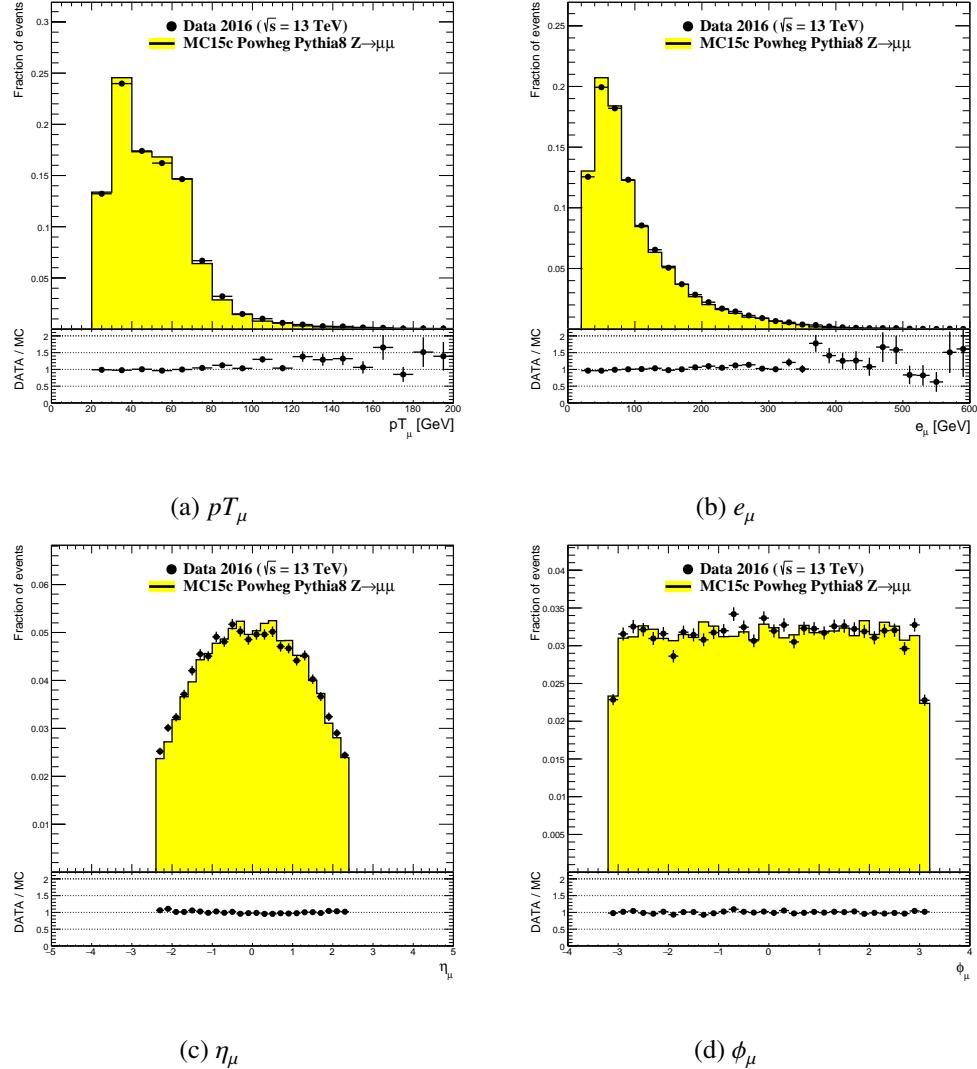


Figure 5.2: Properties of the two muons. The histograms are normalized. The distributions shown are: (a) the transversal momentum of the muons; (b) the muons' energy; (c) the muons'  $\eta$  and (d) the muons'  $\phi$

### 5.4 Reconstruction of the Z-Boson

The Z-boson is reconstructed from the two muons in the selected event. It is required to be in an mass range of  $(90 \pm 10)$  GeV and to have a transversal momentum greater than 30 GeV.

Figure 5.3 shows the properties of the reconstructed Z boson in data-MC comparison. The agreement of data and Monte Carlo for the Z is in general very good. The momentum resolution worsens slightly with  $pT$  exceeding 80 GeV but keeps a reasonable agreement. The mass of the Z boson fluctuates in a range of about 5 GeV around the expected value of 91.19 GeV and has a good agreement with the MC. The agreement of the pseudo-rapidity as well as for the planar angle is also very good.

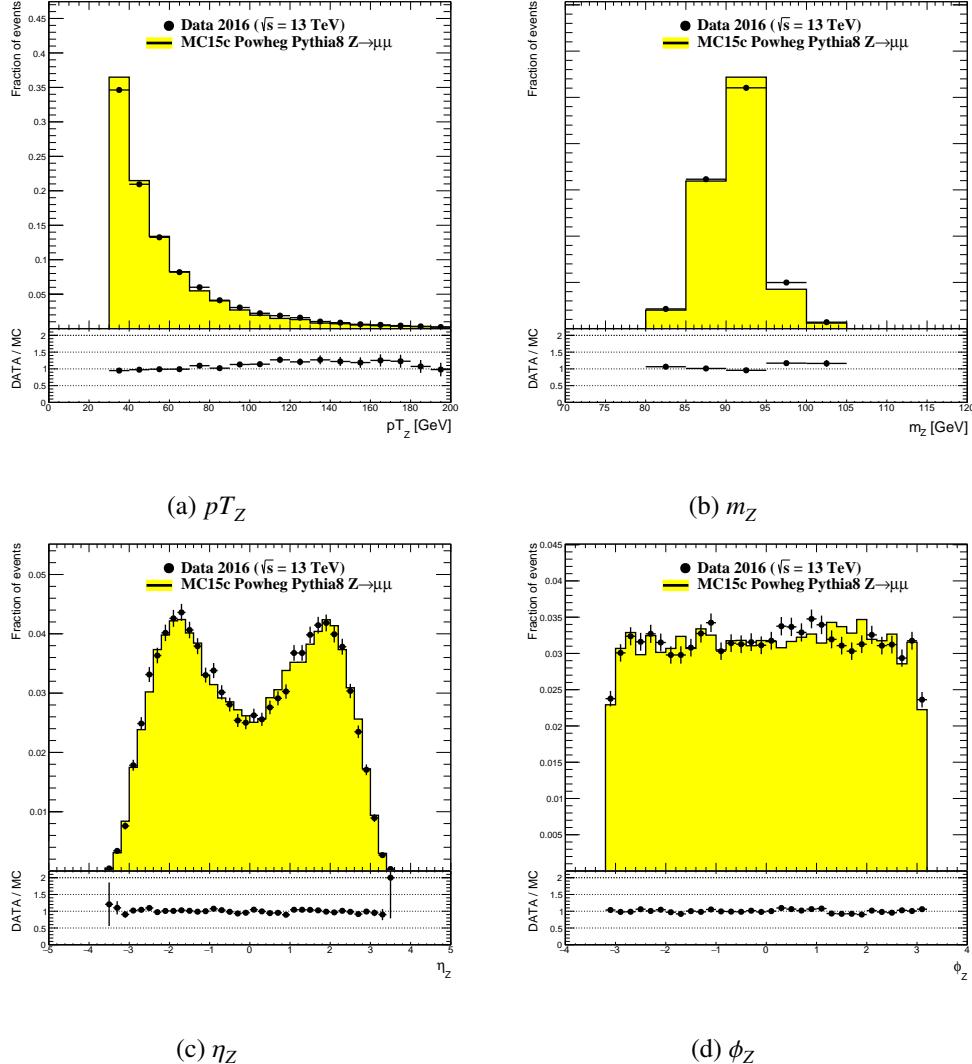


Figure 5.3: Properties of the reconstructed Z-Boson. MC and data are normalized to the total number of reconstructed Z bosons. The distributions show (a) the transversal momentum; (b) the mass; (c) the pseudo-rapidity and (d) the  $\phi$  of the Z.

## 5.5 Kinematic variables of the recoiling jets

To select the jets recoiling to the Z bosons a further selection is applied to a given  $Z \rightarrow \mu\mu$  event. The criteria are as follows:

- The jet is required to have a transverse momentum greater than 20 GeV.

- The jet is required to be recoiling to the reconstructed Z. The criteria is  $|\phi_{jet} - \phi_Z| < (\pi - 0.4)$ .
- Due to the limited region of the tracking detector the pseudo-rapidity of the jet is restricted to  $|\eta| < 2.5$ .

The kinematic properties of the recoiling jet are shown in figure 5.4. The agreement for  $pT$  and energy are rather bad for the jets especially in high energy regions compared to the agreement seen for muons and the reconstructed Z and even in low  $pT$  and energy regions there is a clearly noticeable difference between the two distributions. But in general the shapes for both distributions are well comparable and there is no complete disagreement. The results are close enough to let a good agreement to be expected after further improvements in calibration and scaling. The angular agreement looks better except for  $|\eta| > 2.5$  which is outside the tracker region and could be excluded for Particle Flow performance analysis.

Figure 5.5 shows the comparison of data and Monte Carlo for some further jet variables.

The first distribution 5.5(a) shows the fraction of jet- $pT$  carried by reconstructed tracks. The agreement is decent. The first bin here is not an overflow bin but refers to neutral objects. There are going to be further studies needed to find the origin of this immense fraction of neutral objects.

The second distribution 5.5(b) shows the number of tracks in a jet. The agreement worsens for a high track multiplicity. This might be due to insufficient statistics. Furthermore there might be wrong tracks reconstructed in data because the data multiplicity generally is higher than in MC.

The last distribution 5.5(c) shows the track width. The agreement is reasonable although the fluctuations can hopefully be mediated by further scaling.

## 5.6 Conclusion of the performance

All things considered the data/MC comparison shows that the Particle Flow framework is on a good way. The general shapes of all distributions in our study match well and there are no signs of major problems.

Right now a full jet calibration as well as jet cleaning for Particle Flow jets is not available. Furthermore the scale factors for all objects and further re-weighting for MC and data have to be applied. With the improvements expected from these additional tools it is very likely that similar or better results as in Run 1 can be achieved using the Particle Flow algorithm.

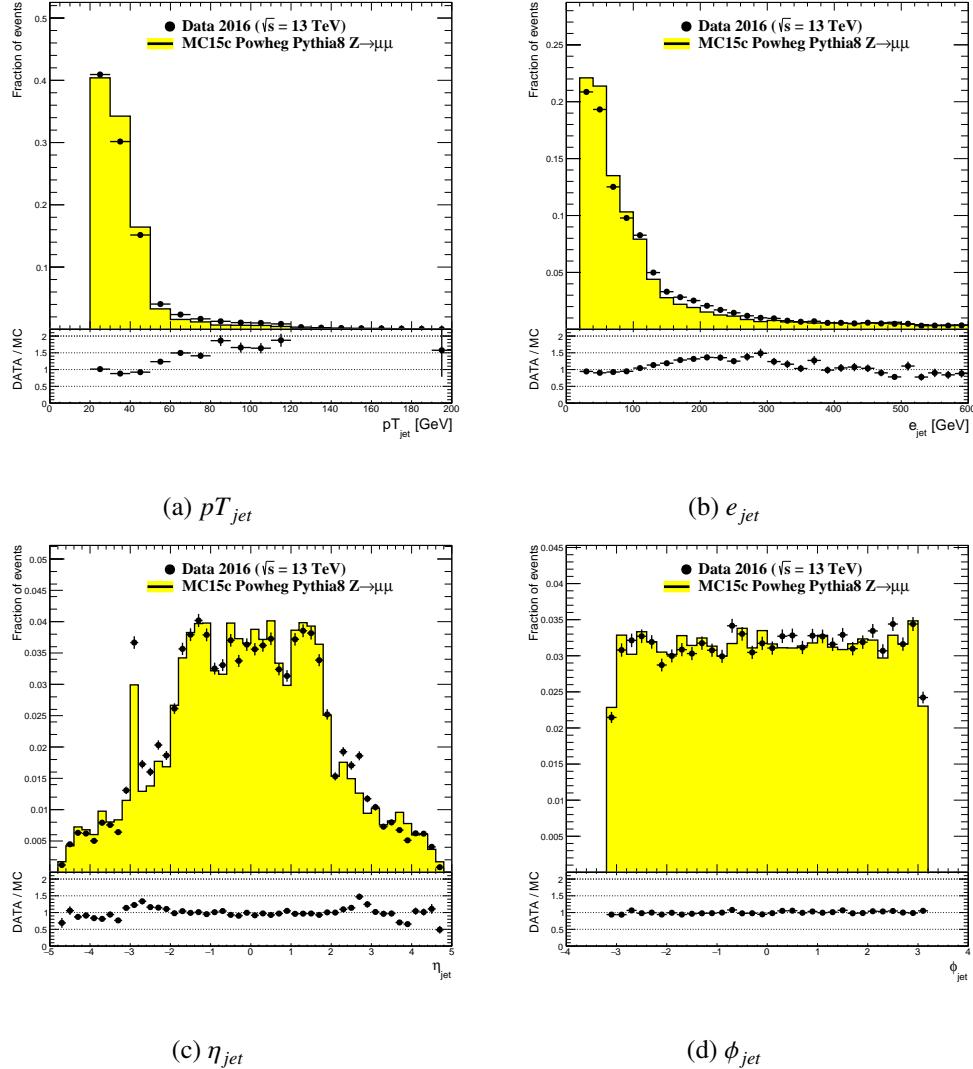


Figure 5.4: Properties of the recoiling jet. The distributions are normalized to the number of jets. Distributions shown are: (a) the jet's transversal momentum; (b) the jet's energy; (c) the jet's eta and (d) the jet's phi.

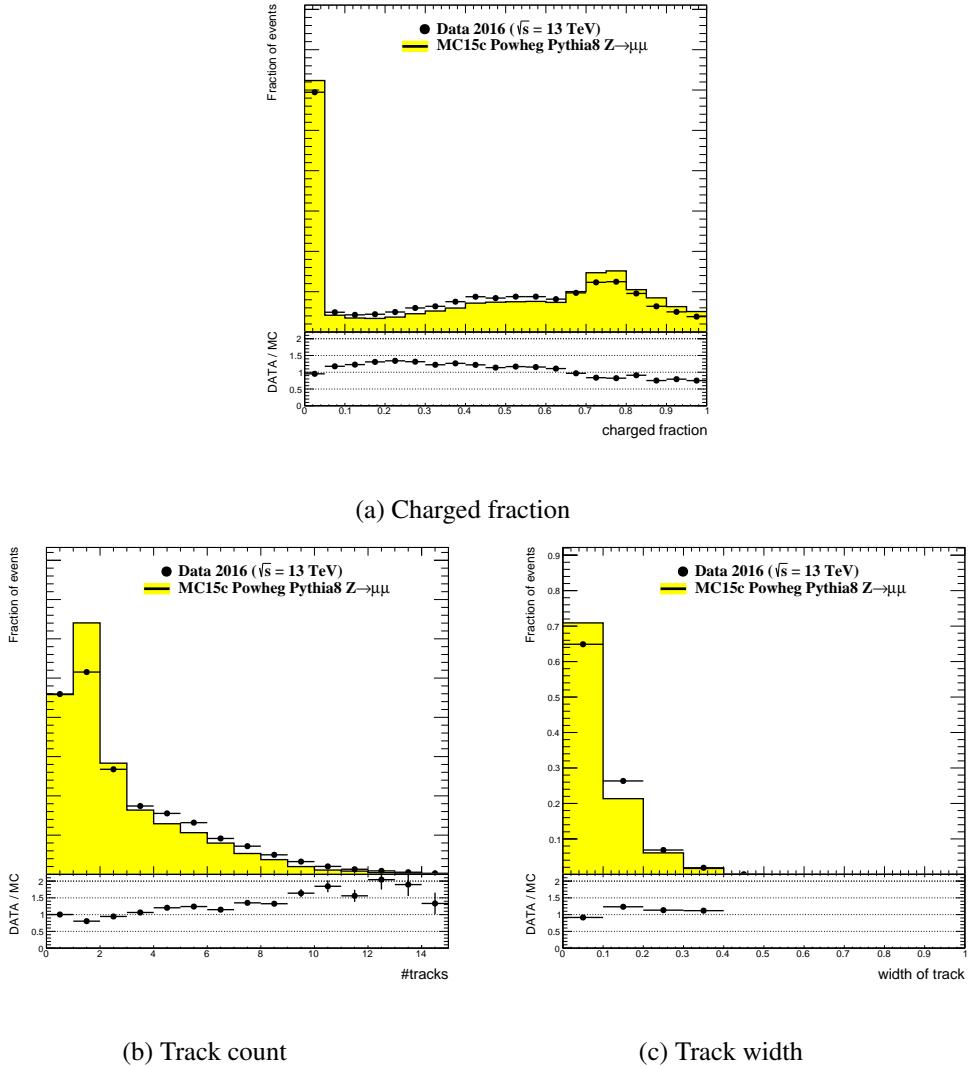


Figure 5.5: Further properties of the recoiling jet in data/MC comparison. The distributions are normalized to the sample size. The following distributions are shown: (a) the charged fraction, i.e. the fractional jet  $p_T$  carried by reconstructed tracks; (b) the number of tracks in the jet; (c) the track's width.

## **APPENDIX A**

---

### **Useful information**

---

In the appendix you usually include extra information that should be documented in your thesis, but not interrupt the flow.



# List of Figures

---

2.1	Sketch of the LHC ring, the position of the experiments and the surrounding countryside.	5
2.2	Sketch of the ATLAS detector	6
2.3	Sketch of the transversal section of the ATLAS detector	7
3.1	Flowchart of the steps of the Particle Flow algorithm	12
3.2	Example of cell subtraction. In red the energy deposited by the $\pi^+$ of interest are shown and in green a cluster from a $\pi^0$ is shown. The algorithm successfully determines the cells belonging to the $\pi^+$ and removes them while leaving the green cells as remnants. Only in subfigure 3.2(f) part of the green cells is removed while part of the red cells remain because both fall into the same subtraction ring. [pflow16]	14
4.1	Fraction of events removed due to the GRL, trigger, trigger matching and event selection	18
4.2	Jet Calibration and Smearing	20
4.3	Calibration and smearing of the muon momentum. The changes are very minimal.	21
4.4	Calibration and smearing of the electron momentum.	22
5.1	Decay of a Z-Boson to two muons	23
5.2	Properties of the two muons. The histograms are normalized. The distributions shown are: (a) the transversal momentum of the muons; (b) the muons' energy; (c) the muons' $\eta$ and (d) the muons' $\phi$	24
5.3	Properties of the reconstructed Z-Boson. MC and data are normalized to the total number of reconstructed Z bosons. The distributions show (a) the transversal momentum; (b) the mass; (c) the pseudo-rapidity and (d) the $\phi$ of the Z.	25
5.4	Properties of the recoiling jet. The distributions are normalized to the number of jets. Distributions shown are: (a) the jet's transversal momentum; (b) the jet's energy; (c) the jet's eta and (d) the jet's phi.	27
5.5	Further properties of the recoiling jet in data/MC comparison. The distributions are normalized to the sample size. The following distributions are shown: (a) the charged fraction, i.e. the fractional jet $p_T$ carried by reconstructed tracks; (b) the number of tracks in the jet; (c) the track's width.	28



## List of Tables

---

2.1	Lepton properties	4
2.2	Quark properties	4
2.3	Mediator properties	4