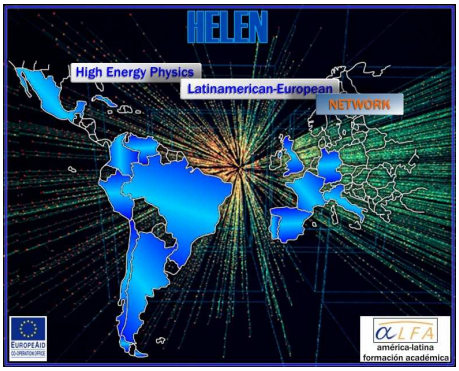

High Energy Physics LatinAmerican-European Network

Fellowship done at
Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE)
CNRS/IN2P3, Université Pierre et Marie Curie-Paris 6,
Université Denis Diderot-Paris 7

Development of an analysis chain for *b*-tagging studies in the ATLAS experiment

Heberth Torres
Universidad de Los Andes, Mérida, Venezuela
Director of internship : F. Derue (LPNHE)
2009

Abstract

Many of the interesting physics processes that will be studied at the ATLAS experiment include the presence of *b*-quarks. During my internship at the LPNHE, I was collaborating in the development of an analysis chain to identified jets of particles coming from *b*-quarks. The method used tags jets by looking for the presence of an electron coming from the semileptonic decay of a *B*-hadron. After learning different analysis tools and understanding the data information, a new ATLAS analysis code for *b*-tagging was developed. In the same way, the performance of the tagger algorithm was measure using two different samples of simulated events. It was found a *b*-tagging efficiency of about 7% (including the branching ratio for the decay $b \rightarrow e\nu X$, of $\sim 20\%$) for a rejection of light jets of about 140.

Contents

1 Introduction

HELEN (High Energy Physics LatinAmerican-European Network) is an academic cooperation program between European Union and Latin America. Its aims are "to form young generations of physicists in High Energy Physics", taking advantage of the european experiments in this physics field; such as the ones at the LHC (Large Hadron Collider). And "to facilitate access of Latin American countries to the technological benefits in the accelerator, detector and information technology domains" [1], between others.

The *Universidad de Los Andes* in Mérida (Venezuela) takes part of the HELEN network. As a student of this university, I had the opportunity of doing a HELEN internship during eight month, at the LPNHE (*Laboratoire de Physique Nucléaire et de Hautes Energies*). The LPNHE, sited in the university campus Jussieu in Paris (France), is a laboratory of the CNRS, IN2P3 and Universities Paris 6 and Paris 7. The activities of this intership were guided by Frederic Derue and took part in the activities of the ATLAS collaboration group at this laboratory.

ATLAS is one of the detectors, experiments, built at the Large Hadron Collider (LHC) at CERN. It is designed to put in evidence the existence of the Higgs boson, to measure more precisely Standard Model parameters and to search signatures of new physics phenomena, among some others purposes [2]. Now that the detector is already built, the ATLAS experiment is in a period of understanding the way how this machine works, learning to identify detected particles and developing and improving tools for physics analysis.

At the LPNHE, Frederic Derue has been working, since some years ago, in the reconstruction and identification of electrons [3, 4] and, using electrons, he is currently working in the identification and labeling of the jets of particles coming from *b*-quarks [5]. The identification of *b*-jets is very important in the ATLAS experiment, since it is required in the analysis of a wide variety of interesting physics processes. During my internship, I was working with Frederic in the development of an analysis chain for *b*-tagging studies, particularly oriented to study the soft electron *b*-tagging method.

In this report, it could be found a brief overview of the Standard Model of Particle Physics in section 2, and a description of the ATLAS experiment in section 3. Next, a brief analysis of Z^0 events is shown in section 4, mainly, including the reconstruction and identification of their electronic decays; an analysis that was done at the beginning of the internship in order to learn several tools and familiarize me with the ATLAS framework. The section 5 contents a description of the jet reconstruction algorithm and presents an analysis over the jets samples. Finally, the section 6 includes the soft electron *b*-tagging method, as well as the results obtained in the measurement of the algorithm performance.

2 The Standard Model of particle physics

Particle Physics include all the research aimed to increase our knowledge about quarks and leptons (which are considered so far the elementary constituents of matter) and their fundamentals interactions. The Standard Model (SM) is a quantum theory that summarizes the current knowledge in this physics field [6]. It describes the composition of matter with twelve elementary constituents and successfully explains three of the four fundamentals interactions, electromagnetic, weak and strong.

The elementary particles constituents of matter are fermions. They have half integer spin and do obey the Pauli exclusion principle, which means that two of them can not be in the same quantum state and the same place simultaneously.

There are two types of fermions, **leptons** and **quarks**. On the one hand, leptons, such as the electron, have integer electric charge (considering the proton's charge as the unit), they do not experience strong interaction and they are able to exist as free particles. On the other hand, quarks have fractional electric charge, experience all four fundamental interactions and can not be isolated.

Quarks and leptons are organized in three families, as it is shown in Table 1. Each family is composed by a lepton negatively charged, a neutrino and two quarks, one quark has electric charge $2/3$ and the other one $-1/3$.

For every particles type there is a corresponding antiparticles type, denoted by a bar over the particles symbol. Particles and antiparticles have identical mass and spin but opposite charges.

	1^{st} family	2^{nd} family	3^{rd} family
Quarks	$\begin{pmatrix} up\ u \\ down\ d \end{pmatrix}$	$\begin{pmatrix} charm\ c \\ strange\ s \end{pmatrix}$	$\begin{pmatrix} top\ t \\ bottom\ b \end{pmatrix}$
Leptons	$\begin{pmatrix} neutrino\ \nu_e \\ electron\ e \end{pmatrix}$	$\begin{pmatrix} neutrino\ \nu_\mu \\ muon\ \mu \end{pmatrix}$	$\begin{pmatrix} neutrino\ \nu_\tau \\ tau\ \tau \end{pmatrix}$

Table 1: Fermions particles of the Standard Model.

According to the Standard Model, **the fundamentals interactions** between the matter constituents are the result of an exchange of particles carriers of force. These particles are **gauge bosons**. Bosons have integer spin and do not obey the Pauli exclusion principle. They are the photons for the electromagnetic interaction, the vector bosons $W^{\pm}$ and Z^0 for the weak interaction, and eight gluons for the strong one.

An important success of the SM is that the electromagnetic and weak interactions have been combined into a unified electroweak theory.

Quarks and gluons are the uniques particles carriers of strong charge (also called "color charge") so they are the only ones that can have strong interactions. They cannot be isolated, but they are confined in color-neutral particles called **hadrons**. Examples of hadrons are the protons and neutrons, the constituents of the atoms nucleus. Two types of hadrons have been observed in nature, **mesons** composed by a quark and an antiquark, and **baryons** composed by three quarks, like protons (uud) and neutrons (udd).

The SM also predicts the existence of an additional particle that has not been observed yet, it is the **Higgs boson**. This one is considered the responsible for the diverse mass values of the gauge bosons and fermions. The reason why it has not been observed is attributed to its hypothetical large mass.

The Standard Model has been consistent with the experimental observations so far and it answers many of the questions in the particle physic field. Nevertheless, there are still several unanswered questions! [6, 7] Some of them are the following:

- How can we unify the electroweak interaction with the strong and gravitational ones into a single unified theory?

- At the beginning, the universe contained matter and antimatter in equal amounts. Why did the antimatter disappear? This mystery could be solved thanks to researches carried out on discrete CP symmetry (charge conjugation and parity symmetries) violation.
- Why and how do elementary particles have mass? Discovering the predicted Higgs boson might help to understand that.
- Does the dark matter consist of new types of particles that interact very weakly with ordinary matter?

Physicist are developing news theories and experiments in order to solve these mysteries. SuperSymmetry is one of the most popular extensions of the SM. It predicts the existence of the so-called supersymmetric particles that are counterparts of the particles which constitute ordinary matter. Then, searching new particles, such as the supersymmetric particles, is one of the mains tasks for the news experiments in Particle Physics.

3 The ATLAS experiment

ATLAS is one of the detectors experiments built at the **Large Hadron Collider** (LHC); the world's newest and most powerful tool for Particle Physics research.

3.1 The Large Hadron Collider

The LHC is a two-ring-superconducting-hadron accelerator and collider constructed at the European Laboratory for Particle Physics (CERN). It has been designed to collide protons with a centre-of-mass energy of 14 TeV and a luminosity[1)]of 10^{34} $cm^{-2}s^{-1}$. These conditions have never been achieved before in any experiment. As well, the LHC will collide heavy ions, specifically lead nucleus, with an energy of 2.8 TeV per nucleon and a peak luminosity of 10^{27} $cm^{-2}s^{-1}$ [8].

In this experiment, protons, before been injected into the LHC, will be progressively accelerated through a set of linear and circular accelerators (see left side of Fig. 1). Protons will achieve 50 MeV of energy in a linear accelerator (Linac), then 1.4 GeV in the Synchrotron Booster (PSB), 25 GeV in the Proton Synchrotron (PS) and 450 GeV in the Super Proton Synchrotron (SPS) [7]. Then, they will be injected into the two mains LHC rings, assembled in bunches with around 10^{11} protons per bunch, in both directions, clockwise and anti-clockwise. Once there, these protons bunches will be accelerated up to 7 TeV (energy per proton) and finally collided at four different points where huge detectors have been constructed to probe the showers of particles product of such collisions. The bunch crossing rate, at each of these points, will be about 40MHz.

There are six detectors installed at the LHC; ATLAS, CMS, ALICE, LHCb, TOTEM and LHCf. Four of them are shown in the right side of Fig. 1. ATLAS and CMS are designed to

[1)] Luminosity L is an important quantity to characterize the performance of an accelerator as it relates the cross-section σ of a given process to the corresponding event rate R:

$$R = L \times \sigma \tag{1}$$

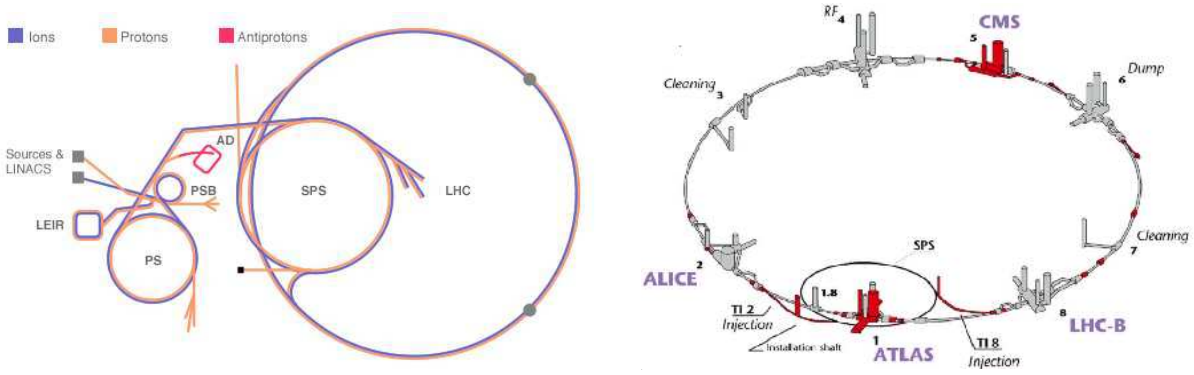

Figure 1: Left: Sketch representative of the accelerators complex at CERN [7]. Right: Sketch representative of the LHC and four of the experiments installed there, ATLAS, CMS, ALICE and LHCb [7].

cover the widest possible range of physics in proton-proton collisions, while LHCb and ALICE are designed to study specific phenomena, LHCb for B-physics and ALICE for the interactions in heavy ions collisions. The detectors used by the TOTEM and LHCf experiments are positioned near the CMS and ATLAS detectors, respectively. TOTEM and LHCf are designed to focus on particles which are scattered forward.

3.2 The ATLAS experiment

ATLAS is the largest-volume detector ever constructed, its dimensions are roughly: 44 m long, 25 m high and 25 m wide, and weight 7000 tonnes [2]. The main feature of this detector is its enormous toroidal-shape magnet system (see Fig. 2), and that is why it is called ATLAS, A Toroidal LHC ApparatuS. The toroidal magnet consists of eight 25 m long superconducting magnet coils, arranged to form a cylinder around the beam pipe through the centre of the detector.

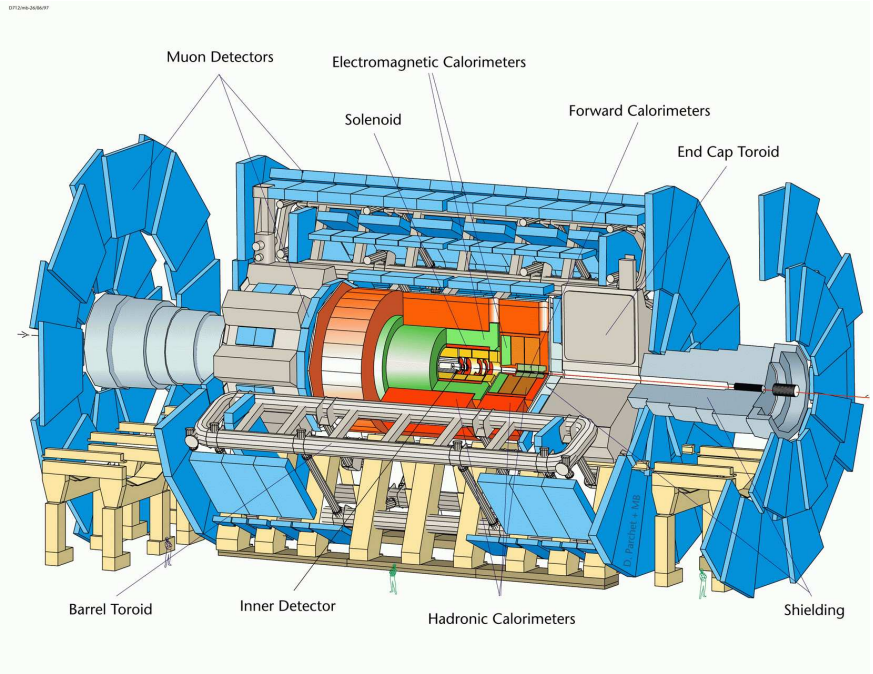

Figure 2: Schematic view of the ATLAS detector with its different constituents parts [9].

This experiment is the product of an international collaboration, in which participate over two thousand physicist and engineers, from 169 institutes, from 37 countries.

The detector is constituted by four sub-detectors; these are the inner detector, the electromagnetic calorimeter, the hadronic calorimeter, and the muons spectrometer. A sketch representative of different kind of particles passing through the ATLAS sub-detectors is shown in Fig. 3. The inner detector measures the tracks of charged particles which are bent by the magnetic field of a superconducting solenoidal magnet. Then, outside of the solenoid, the two calorimeters measure the energies of particles. Finally, the muons spectrometer measures the tracks of muons which are bent in the field of the toroidal magnets.

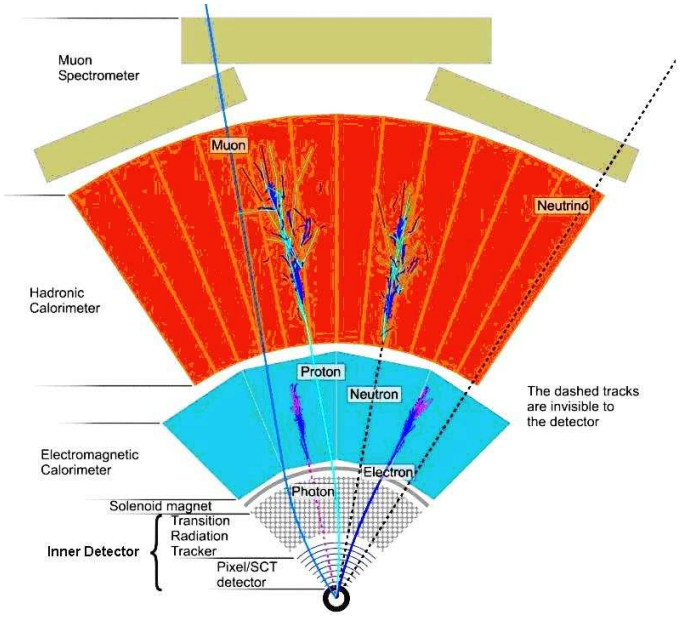

Figure 3: Sketch representative of different kind of particles passing through the ATLAS sub-detectors [9].

3.3 Coordinates system and some important variables

The coordinates system employed to describe the ATLAS detector and the motion of particles passing trough it is shown in the left side of Fig. 4. The nominals interaction point is defined as its origin, the beam direction sets the *z*-axis, and the *x*-*y* plane is the transverse plane to the beam direction. The positive *x*-axis is defined as pointing from the collision point to the centre of the LHC ring and the positive *y*-axis is pointing upwards. The azimuthal angle ϕ is measured around the beam axis, and the polar angle θ is the angle from the beam axis [2].

There are also some others variables frequently employed. They are:

- The pseudorapidity η, defined as

$$\eta = -\ln(\tan(\theta/2)). \tag{2}$$

  This angular quantity is more commonly utilized as polar coordinate than θ. Their correspondence for some specifics angles is presented in the right side of Fig. 4.

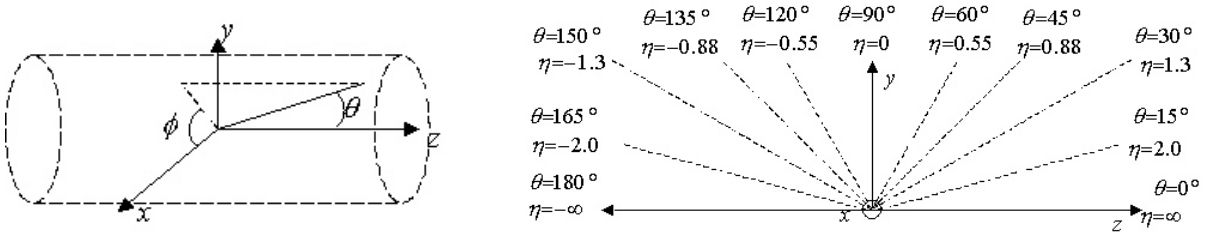

Figure 4: Left: Coordinate system used in ATLAS [7]. Right: Correspondence between η and θ for some specifics values [7].

- The distance ΔR in the pseudorapidity-azimuthal angle space, defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{3}$$

- The transverse momentum p_T defined in the $x-y$ plane.
- The transverse d_0 and the longitudinal z_0 impact parameter, defined for a particle track as the transverse and the longitudinal distances to the nominal collision point at the point of closest approach.

3.4 The ATLAS detector components

3.4.1 The Inner Detector

A schematic view of the Inner Detector (ID) is shown in the left side of Fig 5. It is 6.2 m long and it has a 2.1 m diameter. The ID has as task to measure the track of charged particles; for this, it has a set of layers where is measured the position of the charged particles as they pass throughout each one. Beginning from innermost part, the inner detector has three layers of silicon pixels detectors, four double layers of semiconductor trackers (SCT), and a transition radiation tracker (TRT) (see right side of Fig. 5). All this tracking system is inside a solenoidal magnet that produces a magnetic field of 2 T; thus, charged particles are bent permitting to know their charge and their momentum.

The precision tracking detectors (pixels and SCT) cover the region $|\eta| < 2.5$. In the barrel region, they are arranged on concentric cylinders around the beam axis; while in the end-cap regions, they are located on disks perpendicular to the beam axis, such as it is shown in the left side of Fig. 5. The silicon pixel detectors have a very high granularity and they are very close to the beam. The minimum pixel size is $50 \times 400\mu\text{m}^2$ ($R-\phi \times z$) and the first one of these pixels layers is at just 5 cm from the beam. These pixels layers with such characteristics together to the SCT layers permit to measure with high performance the impact parameter (cf. section 3.3). Thus, it enables to determine whether a particle was originated at the collision point (primary vertex) or a few millimeters from it as a decay product of another particle (secondary vertex).

The transition radiation tracker (TRT) covers the region $|\eta| < 2.0$. It only provides two-dimensional measurements, $R-\phi$, but it gives a large number of these ones per track, typically 36. The TRT consists of several tens of thousands, 4 mm diameter, straw tubes. They are parallels to the beam axis in the barrel region and they are arranged radially in wheels in the end-cap region. These straws are filled with a gas that is ionized whenever charged particles pass through. Each straw has in the center a wire that is maintained at a potential higher than

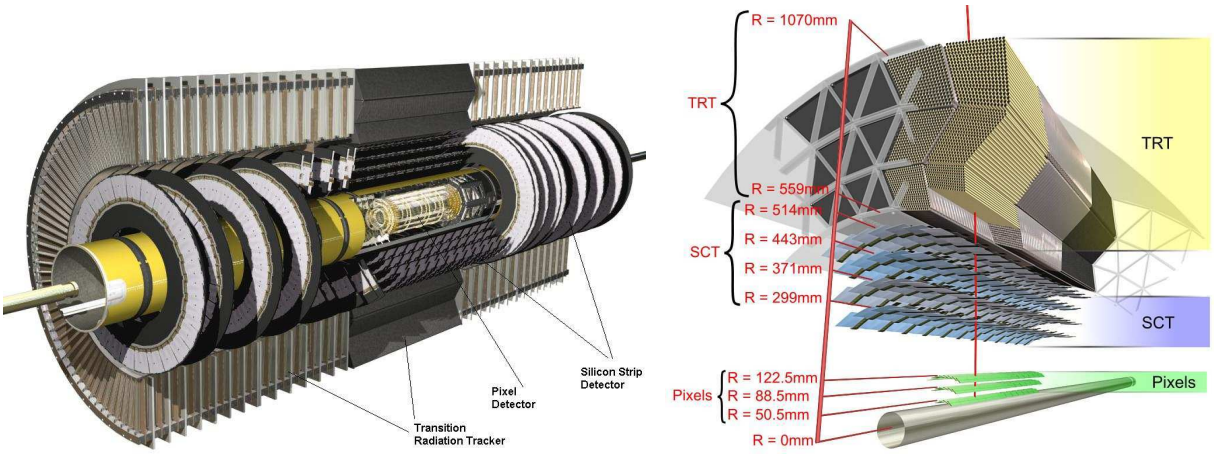

Figure 5: Left: Schematic view of the Inner Detector [9]. Right: Structure of the inner detector in the barrel region [9].

the inner wall. So, the ions are drawn to the inner wall, while the electrons are drawn to the central wire causing to appear a current through this wire.

The straws also contain materials with different indices of refraction. Then, when the charged particles pass through, they radiate photons that also ionize the gas and therefore they become the signal stronger. Because electrons radiate more photons than some others particles, like pions, the TRT has also electron identification capabilities.

3.4.2 The Calorimeters

The ATLAS calorimeters are presented in Fig. 6. Electromagnetic calorimeters cover the region $|\eta| < 3.2$, hadronic calorimeters cover the same region and forward calorimeters cover the region $3.1 < |\eta| < 4.9$. In these calorimeters, the energy of particles is measured by stopping them with dense materials, there the particles interact generating large showers of secondary particles before been completely stopped.

The electromagnetic calorimeter has as main objective to measure the energy and the motion direction of electrons and photons. It is divided into a barrel part ($|\eta| < 1.475$) and two end-cap components ($1.375 < |\eta| < 3.2$); all these parts work in a similar way. This calorimeter is an accordion-shape structure that consists of many layers of lead electrodes and liquid argon (LAr). There is also a copper grid immersed in each liquid argon layer that acts as an electrode. In this case, the particles interactions with the lead plates generate electromagnetic showers. Then, all these electromagnetic particles ionize the argon as they pass through. The electrons resulting from the ionization are drifted to the copper electrodes and the electric current is measured. The greater is the energy of a particle entering in the EM calorimeter, the greater will be the shower generated, and the current in consequence. The accordion geometry of this calorimeter provides complete ϕ symmetry without azimuthal cracks.

In front of the calorimeter, there is also a LAr layer with an electrode. The information that it provides is utilized to correct the energy lost by electrons and photons when they go through the matter in front of the calorimeter.

The hadronic calorimeter measures the energy of particles that traverse the EM calorimeter with almost not interaction; these are primarily hadrons. This calorimeter is divided into a

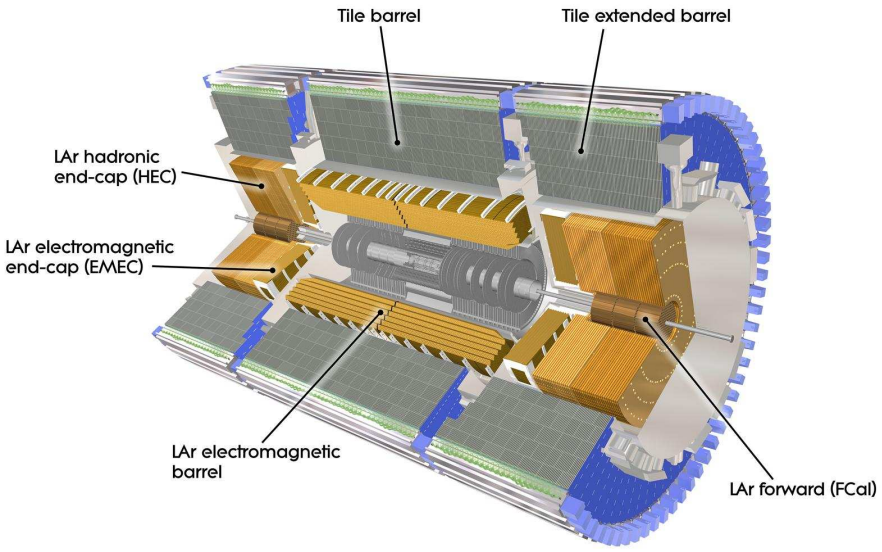

Figure 6: Schematic view of the ATLAS calorimeters [9].

barrel part ($|\eta| < 1.7$) and two end-cap components ($1.7 < |\eta| < 3.2$) as well. But in this case, these parts work differently; the tile barrel calorimeters utilize scintillating plates and the end-caps are liquid argon calorimeters. That is because the radiation emanating from the collision point is more intense at large values of η, and the scintillating tiles are damaged by excessive exposure to radiation.

The tile barrel calorimeter utilizes steel sheets in order to generate the hadronic shower and scintillating sheets as the active material. They are placed in planes perpendiculars to the beam, forming layers of steel and the scintillating material. When the shower particles pass through the scintillating tiles, they make them emit light in an amount proportional to the incident energy. Then fibers carry the light to devices where the light intensity is measured.

The liquid argon end-cap hadronic calorimeter is very similar to the EM calorimeter. The difference is that it uses copper plates instead of lead plates, which are more appropriate to the hadronic showering process, and the argon gaps are twice larger as well.

The forward calorimeter consists of three modules in each end-cap. The first one, made of copper, is optimized for electromagnetic measurements, while the other two, made of tungsten, measure predominantly the energy of hadronic interactions. In this calorimeter, the liquid argon is again utilized as the active material.

3.4.3 The Muon Spectrometer

A view of the muon spectrometer is shown in Fig. 7. This sub-detector measures the muons tracks and their momentum using the deflection caused by the superconducting toroidal magnets. There are three toroidal magnets; the large barrel toroid and two smaller end-cap magnets, which are inserted into both ends of the large one. In the barrel region, tracks are measured in chambers arranged in three cylindrical layers around the beam axis, while in the end-cap region, the chambers are installed in planes perpendicular to the beam, also in three layers. In

this muon detector several types of chambers are used for different purposes.

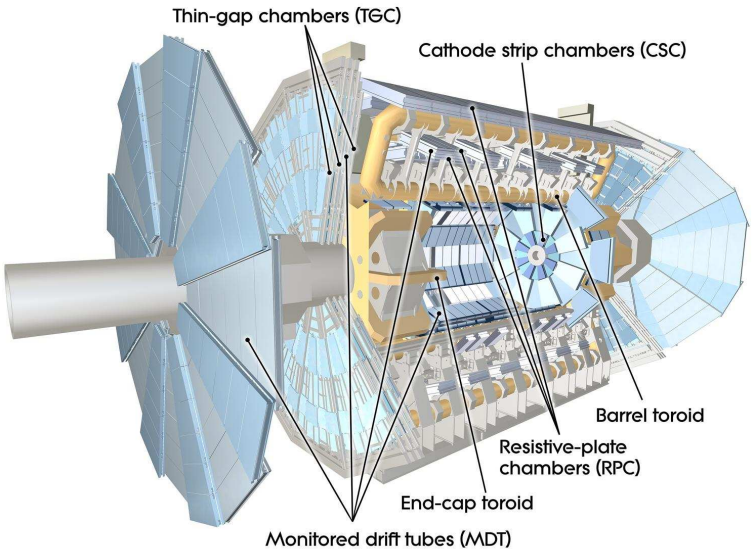

Figure 7: Schematic view of the ATLAS muon spectrometer [9].

On the one hand, in order to make precision measurements of the track coordinate in the principal bending direction, the chambers contain Monitored Drift Tubes (MDT). They are used in two regions, the barrel and the external part of the end-caps ($|\eta| < 2.0$). In the internal part of the end-caps ($2.0 < |\eta| < 2.7$), it is necessary higher granularity. Therefore, the chambers in this region are of another type, they are Cathode Strip Chambers (CSC).

On the other hand, in order to provide information to the trigger system, Resistive Plate Chambers (RPC) are used in the barrel region, and Thin Gap Chambers (TGC) are used in the end-caps region. The purposes of these trigger chambers are: to provide bunch-crossing identification as well as well-defined p_T thresholds for the trigger, and to measure the muon coordinate in the direction orthogonal to that determined by the precision-tracking chambers [2].

3.5 The trigger system

The trigger system is basically a filter that has as task the selection of the interesting events. It has to decrease the event rate from the brunch crossing rate of 40MHz to a rate of about 100Hz [10], a rate at which data can be written to permanent storage. The trigger system consists of three different levels, level 1 (L1), level 2 (L2) and an event filter (EF).

The first level uses a limited amount of the total detector information to take a decision in less than 2.5 μs, and it reduces the rate to about 75 kHz. The L1 trigger searches for high transverse-momentum muons, electrons, photons, jets, and τ-leptons decaying into hadrons, as well as large missing and total transverse energy. In order to do that, L1 utilizes the trigger chambers in the muon spectrometer, and reduced-granularity information from all the calorimeters. For each event, the L1 trigger also defines one or more Regions-of-Interest (RoIs); i.e. the geographical coordinates in η and ϕ of those regions in the detector where its selection process has identified interesting features. This information is then used by the higher levels trigger.

The second level uses, at full granularity and precision, all the information available within the RoIs (approximately 2% of the total event data). The L2 is designed to reduce the trigger rate to approximately 3.5kHz, with an event processing time of about 40 ms. Finally, in the EF trigger, events are reconstructed completely and some offline analysis procedures are used to make the selection. It is done in about four seconds and it reduces the event rate to roughly 200Hz, such as it is desired.

All the selection criteria used by the different trigger levels are organized in trigger menus. These trigger menus serve to classified events into physics channels. For example, there is a trigger menu called "$e25i$" that requires at least one isolated electron with more than 25GeV. Then, events that pass through this trigger menu are known to be either type $W \rightarrow e\nu$, $Z^0 \rightarrow e^+e^-$, quark top, $H \rightarrow WW^*/ZZ^*$, W', or Z' [10]. The criteria utilized in these menus can be optimized as the luminosity change.

3.6 Data analysis model

3.6.1 Athena software

In order to provide to physicists the tools necessaries to process and analyze the data coming from the trigger and the acquisition systems, a software framework called Athena will be used by the ATLAS collaboration. It is based on a software architecture called Gaudi that was originally developed by the LHCb experiment. Nowadays the Gaudi project is a kernel of software common to both experiments and co-developed, while Athena is the sum of this kernel plus ATLAS-specific enhancements. This framework is primarily based on the C^{++} programming language and some components implemented use Fortran and Java. Athena has been used until now in the production and analysis of simulated data, and in the analysis of real data from the detector's tests and from cosmic rays that go through the detector [10].

For a period during my internship, I was collaborating in the development of an analysis code for *b*-tagging that uses and takes part of the Athena tools.

3.6.2 Simulations and data model

Since the beginning in the ATLAS experiment, there has been the necessity of doing simulations in order to prepare physics analysis, as well as to evaluate both the detector and the analysis performance. The ATLAS simulation program can be divided into three separate modules; event generation, detector simulation, and digitization. Then, the simulated data pass through two additional stages, the reconstruction and the creation of data files utilized in physics analysis. In Fig. 8 a schematic representation of this steps chain is shown.

In the **event generation**, it is modeled the complex physics processes that lead to the production of hundreds of particles per event at LHC energies. Generators model the initial and final-state radiation in these processes, multiples interactions and beam remnants, as well as hadronization and decays. The event generators most commonly used in ATLAS are Pythia and Herwig; they are based on the Monte Carlo technique. The generators output is converted into a common format called HepMC (Monte Carlo Generator 4-Vector Classes).

The ATLAS **detector simulation** is based on GEANT4; a toolkit which provides both a framework and the necessary functionality for running detector simulations in Particle Physics. It includes optimized solutions for geometry description and navigation through the geometry,

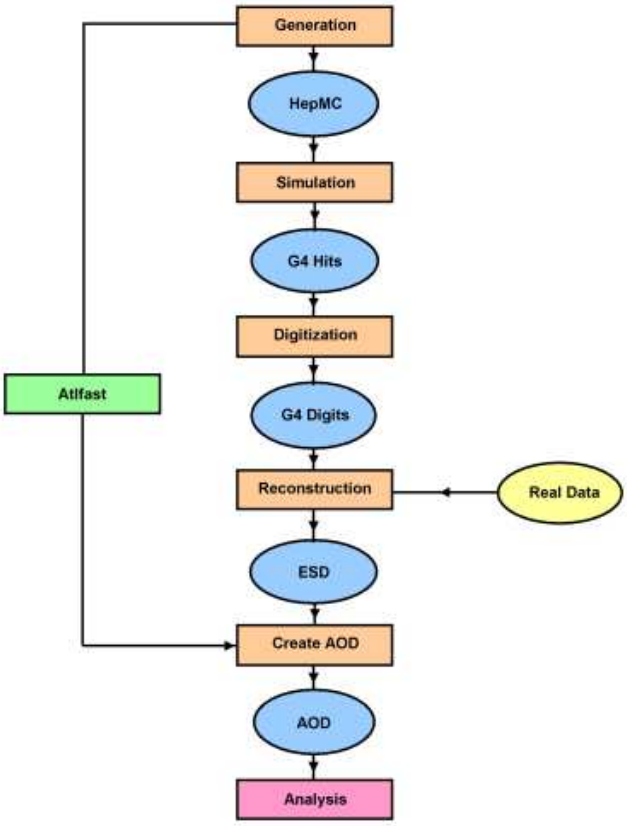

Figure 8: Schematic representation of the full chain of steps in the production of simulated data [11].

the propagation of particles through detectors, description of materials, as well as modeling of physics processes.

Then, the hits produced by G4ATLAS need to be translated into the output actually produced by the ATLAS detector. This is the **digitization**. In this step the propagation of charges or light into the active media as well as the response of the readout electronics are taken into account, for example, charges in the tracking detectors and the liquid argon calorimeter or light in the case of tile calorimeter. The final output of the digitization step are Raw Data Objects (RDO's), with ~ 1.6 MB/event [11]. This data should resemble the real detector data.

Following in the **reconstruction**, matrices containing the energies in all calorimeter cells are filled, and clusters of energy are reconstructed. Muon track segments in the muon system are found from a combinatorial search of the single track segments. Hit coordinates are reconstructed in the precision tracker and in the TRT, and tracks from charged particles are searched for. Then, the information from all the sub-detectors is combined to get the most accurate measurements of objects such as jets, electromagnetic particles, muons, etc. All the detailed information output of the reconstruction is written in a format called Event Summary Data (ESD), with ~ 1 MB/event. There is also a format similar to the ESD one that is directly usable by ROOT; this format is ComBined NTuple (CBNT). ROOT is the analysis and histogram visualization package used in Particle Physics.

For physics analysis is created a summary of the event reconstruction information that contains objects such as electrons, photons, muons, jets, etc. In this case the format is named Ana-

lysis Object Data (AOD), and it has an acceptable size for wide distribution ($\sim$ 100 kB/event). Finally, derived from the AOD format, a Derived Physics Data (DPD) is created to reduce further the size of the analysis objects. These DPDs will be defined by the different physicists according to the analysis that they are doing. On average DPDs will contain 10 kB/event.

Such as it is shown in Fig. 8, one can also short circuit the full chain by using **Atlfast**. It provides a fast simulation of the whole chain by taking the generated events and smearing them to produce AOD directly. Atlfast can in fact take input from either the event generator, simulation, digitization, or ESD files.

3.6.3 The computing model

When the LHC be operating, it will be producing an enormous amount of data, roughly 15 PB of data annually [12]. In order to provide to all the scientists around the world participating in the LHC experiences, the computing resources to access and analyze this data, a novel globally distributed computing model will be used. This is the computing grid. It consists of many computers interconnected sharing resources to provide a global power to process and store data. More than 140 computing centers in 34 countries are already joined participating in the project Worldwide LHC Computing Grid (WLCG).

The data from the LHC experiments will be distributed according to a four-tiered model. A primary backup of the RDOs will be recorded on a tape at CERN, the Tier-0 center of LCG. After an initial processing, such as first-pass reconstruction, calibration and alignment, these data will be distributed to a series of Tier-1 centers. These are large computer centers dispersed around the world, with sufficient storage capacity for about 20% of the RDO data [10]. In Tier-1 centers, the reconstruction process will be done with better calibration and alignment.

The Tier-1 centers will make ESD data available to Tier-2 centers. These laters consist of several collaborating computing facilities, which can store sufficient data and provide adequate computing power for specific analysis tasks (mainly with AODs and DPDs). Individual scientists will access these facilities through Tier-3 computing resources. The Tier-3 may consist either of local clusters in a University Department or even of individual PCs, and which may be allocated to LCG on a regular basis [11, 12].

For the finals analysis done during my internship, I was taking advantage of the computing grid facilities. They are already available for the production and analysis of simulated data.

4 Studies with Z^0 events

The Z^0 events will play an important role in the ATLAS detector calibration during the first period of data taking. The production rate of pairs of electrons from Z^0 decays will be about 3,5 Hz [10]. So, these events will be a large source of electrons useful to study the performance of the reconstruction algorithms. Particularly, they will permit to measure the electrons identification efficiency at the trigger, reconstruction and analysis levels. The Z^0 events will also permit to study the alignment between the inner detector and the electromagnetic calorimeter, as well as to establish an energy scale common to both sub-detectors. In the same way, the precise knowledge of the Z^0 boson mass will serve to inter-calibrate the electromagnetic calorimeter.

In this section, I describe a brief study of the electronic decays of Z bosons that I did at the very beginning of my internship. This analysis consisted in the reconstruction of the Z boson

mass with pairs of electrons that have opposite charges after a process of electrons identification and selection. I did this analysis following a "tutorial" in order to learn to utilize ROOT macros as well as the information contained in data samples after the reconstruction process.

4.1 The $Z^0 \rightarrow e^+e^-$ sample

All samples employed have been produced using the Pythia Monte Carlo events generator and the full ATLAS detector simulation based on GEANT4. These data have been produced for the ATLAS Computing System Commissioning.

For this analysis, I utilized firstly a ComBined NTuple (CBNT) data sample, that only contains one thousand events, all of them are type $Z^0 \rightarrow e^+e^-$. I utilized the CBNT format because it is directly accessible by ROOT, and hence it facilitates learning certain tools. Nevertheless, it is a heavy format and it takes much disc space for analyzing large amount of data. Then, most of the plots (the ones that shown the events characteristics and variables discriminating of electrons) have been redone using the framework based on ATHENA that was developed during this internship, and using the Analysis Object Data (AOD) format. For this later part, I utilized two set of data; a signal sample and a background sample. The signal sample consists of 62,600 events type $Z^0 \rightarrow e^+e^-$, and the background sample, of 250,000 dijets events.

4.2 Study at generation level

In Fig. 9, it is shown the distributions of the transverse momentum p_T (left), and of the pseudorapidity η (right) for different particles at the generation level. The distributions shown correspond to Z particles (dots), electrons from the Z decays (hatched histogram), and background electrons (solid line). In this case, the background electrons are those originated from the photons conversion in the matter. On the one hand, the electrons from Z tend to have higher p_T than the electrons from conversion; therefore, in order to reject the last ones, only electrons with $p_T > 20$ GeV will be considered. On the other hand, the electrons from Z are flatly distributed in all the η range. Only electrons with $\eta < 2.5$ are considered, since this is the region covered by the inner detector.

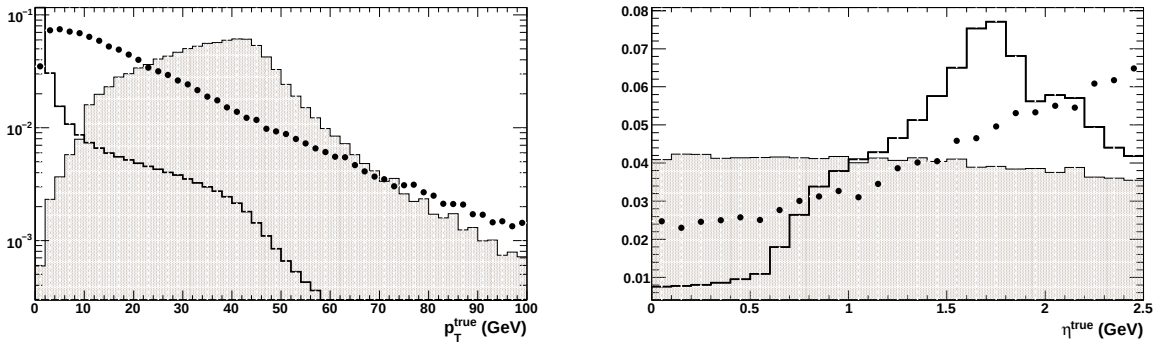

Figure 9: Distribution of transverse momentum p_T (left) and pseudorapidity η (right) for Z particles (dots), electrons from Z (hatched histogram) and background electrons (solid line), at the generator level. These distributions are normalized, and they have been redone after the first original analysis.

4.3 Reconstruction and identification of isolated electrons

For the standard reconstruction of electrons, a seed cluster is taken from the electromagnetic calorimeter, and a loosely matching track is searched for among all reconstructed tracks. Those clusters having matched track are defined as electrons, excluding those which matched track recognized to belong to photon conversion. For all electron candidates, shower-shape variables (lateral and longitudinal shower profiles, etc.) are calculated using the fine granularity of the electromagnetic calorimeter, and typically more than 50 calorimeter cells are summed to collect the full cluster energy. The energy of high-pT (> 20 GeV) electrons is obtained from the energy measured in the calorimeter, while the η and ϕ directions are precisely determined using the associated track.

Then, the standard identification is based on cuts on the shower shapes, on the reconstructed tracks, as well as on combined reconstruction properties — the ratio of energy (calorimeter) to momentum (inner detector), the difference between the coordinates η and ϕ reconstructed by the cluster and the track extrapolated into the calorimeter. These cuts explicitly require the presence of a vertexing-layer (the innermost layer) hit on the track to further reject photon conversions, and a high ratio between high-threshold and low-threshold hits in the TRT detector to further reject the background from charged hadrons. There are about 26 cuts utilized in the standard identification. In my analysis, I just made use of a few of them, and it was mainly oriented to differentiate between the electrons from Z and pions miss reconstructed as electrons, which constitute a large part of the background.

In Fig. 10, two of the electrons discriminating variables that I employed are shown, the hadronic leakage (right) and the ratio E/p of energy to momentum (left). The hatched histograms correspond to electrons from Z, while the histogram with solid line correspond to pions. The hadronic leakage is defined as the ratio between the transverse energy reconstructed in the first compartment of the hadronic calorimeter in a window $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$, and the transverse energy reconstructed in the electromagnetic calorimeter. It is possible to see that electrons deposit only a very small fraction of their energy in the hadronic calorimeter, typically less than 2% according to the figure, while pions deposit much more. The right side of Fig. 10 shows that the energy measured in the EM calorimeter is more consistent with the momentum measured in the tracker for electrons than for pions.

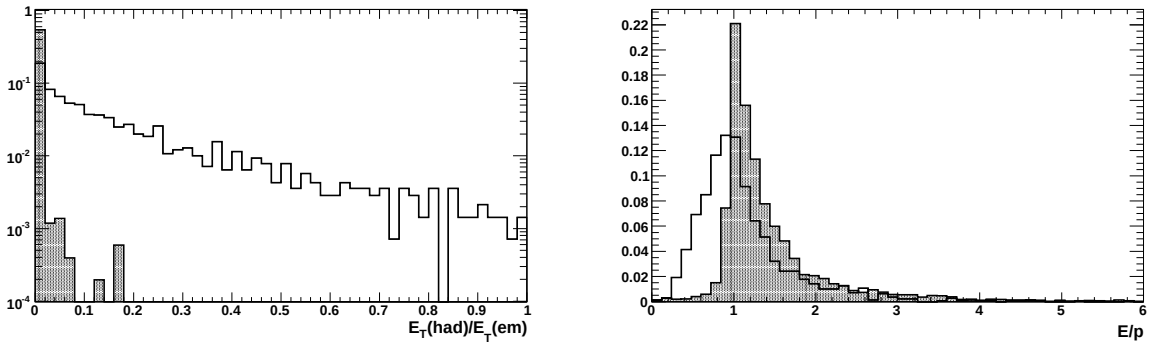

Figure 10: Distributions of the hadronic leakage (left) and the ratio E/p between the energy reconstructed in the calorimeter, and the transverse momentum reconstructed in the inner detector (right). The hatched histograms correspond to electrons from Z while the solid line correspond to pions. These distributions are normalized, and they have been redone after the first original analysis.

In my analysis, after the selection cuts, I achieved an electron from Z selection efficiency of $\sim 60\%$. This efficiency is defined as the fraction of the reconstructed electrons from Z that

have been selected. In Fig. 11 it is shown the selection efficiency as a function of the transverse momentum p_T. The value achieved is the typical value found in the standard isolated electrons identification efficiency.

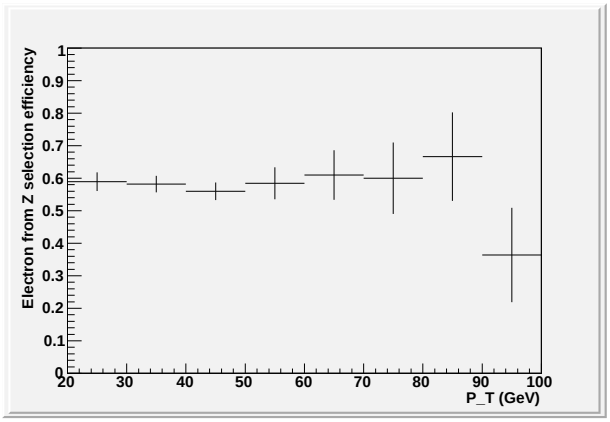

Figure 11: Identification efficiency as a function of the transverse momentum p^T.

4.4 Reconstruction of the Z^0 mass

After the electrons selection, the invariant mass of pairs of electrons with opposite charge was calculated. In Fig 12 it is shown the distribution of this invariant mass. It is roughly consistent with the mass of the Z boson which according to the Particle Data Group is $M_Z = 91.187 GeV/c^2$ [13]. I would like to remain that this analysis was done at the beginning of my internship to simply learn tools, I utilized a sample of only one thousands events, and it was not done any emphasis to improve the results.

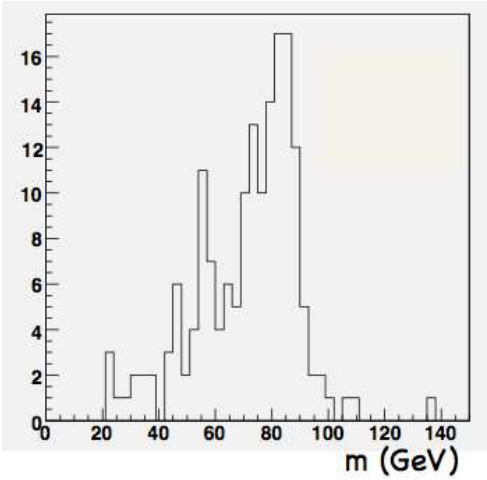

Figure 12: Distribution of the invariant mass of pairs of electrons with opposite charge in the $Z \to e^+e^-$ events.

5 Studies of jets

In high energy collisions, the hadrons and particles resulting from the decay of a high energetic parton (i.e quarks and gluons) remain collimated around the original direction of the parton, such as it is shown in Fig. 13 . The higher is the energy of the parton, more collimated the hadrons will remain. These bunches of hadrons and particles are called jets.

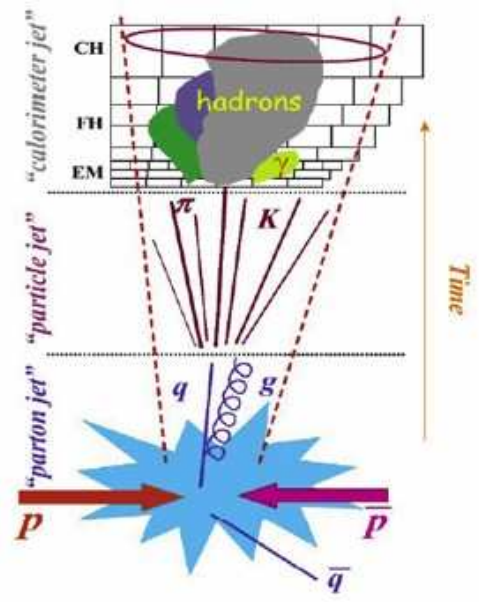

Figure 13: Schematic view of a jet.

At the ATLAS experiment, basically all physics analysis to be performed require the reconstruction of jets with high quality and efficiency. The jets reconstruction in ATLAS is briefly described here and then some results are presented. They were obtained at the beginning of the internship with macros based on CBNT data and then with the new framework developed during the internship.

5.1 Jets reconstruction

Several jets finding algorithms are being studied to produce jets with different qualities and sizes in order to cover the wide variety of physics processes of interest at the LHC. For example, on the one hand for the measurement of the inclusive QCD jet cross-section, wider jets are typically preferred to capture the hard scattered parton kinematics completely. On the other hand, to reconstruct a W boson decaying into two jets or to find jets in very busy final states like $t\bar{t}$ production or possibles SUSY signatures, narrow jets are preferred.

One of the two algorithms most commonly used in ATLAS will be briefly described. This is a seeded fixed cone algorithm. The other one is a successive recombination algorithm; a description of it can be found in [14].

The fixed cone jets finder

The seeded cone algorithm uses two parameters, the transverse momentum threshold for a seed, $p_T = 1$ GeV for all cone jets, and the cone radius R_{cone}, with $R_{cone} = 0.4$ for narrow jets and

$R_{cone} = 0.7$ for wide jets. The finding jets procedure is as follow; first, all the input objects (partons, particles, reconstructed detector objects with four-momentum representation, such as tracks and calorimeter cells) are ordered in decreasing order in transverse momentum. If the object with the highest p_T is above the seed threshold, the next object within a cone in η and ϕ with $\Delta R < R_{cone}$ is combined with the seed, and a new direction is calculated for the cone from the recombination of the two four-momentum. Then the next object within R_{cone} is found and so forth, until the cone is stable, at which point the next seed is taken from the input list and a new cone is formed. This continues until no more seeds are available. When two jets share constituents between them, if they share more than 50% of the p_T of the less energetic jet, they are merged. If the amount of shared p_T is below the 50%, they are split.

Calorimeter jets

In ATLAS, jets are built from the energy reconstructed in the electromagnetic and hadronic calorimeters. The calorimeters system has about two hundred thousand individuals cells of various sizes and with different readout technologies and electrode geometries. For jet finding, it is necessary to first combine these cell signals into larger signal objets with physically meaning four-momentum. The two concepts available are calorimeter *signal towers* and *topological cell clusters*. Only the first will be shortly described.

In the case of the **towers**, the cells are projected onto a fixed grid in pseudorapidity η and azimuth ϕ. The tower bin size is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the whole acceptance region of the calorimeters, with 6400 towers in total. Projective calorimeter cells, which completely fit inside a tower, contribute their total signal to the tower signal. Both non-projective cells and projective cells larger than the tower bin size contribute a fraction of their signal to several towers, depending on the overlap fraction of the cell area with the towers.

In this study, we used jets reconstructed using signals towers, with the fixed cone jets finder and with a cone of 0.4 of radius. In the following part, characteristics found of these jets are shown.

5.2 Jets sample

5.2.1 Jets sample characteristics

As it was said above, all samples used have been generated using the Pythia Monte Carlo events generator and the full ATLAS detector simulation based on GEANT4. These data have been produced for the ATLAS Computing System Commissioning.

One of the samples used in this study contain events with electrons in jets from Higgs boson associated production WH, with $W \to \mu\nu$. In the generation of this samples it has been taken the mass of the Higgs boson to be $m_H = 120$ GeV. The signal sample used consists of 50,000 $H \to b\bar{b}$ events, while the background sample consists 175,000 of $H \to u\bar{u}$ events.

The WH sample has been used in previous studies of b-tagging. But, at beginning period in the LHC, this kind of data will not be available with enough statistics to make analysis. In contrast with this fact, in the same period there will be a large production of QCD di-jets events. That is why in this study we have used simulated samples of di-jets events as well. This last sample consists of 250,000 events.

The jets utilized in the analysis have been preselected according to the standards cuts employed for b-tagging studies in ATLAS [5]. It is required a transverse momentum $p_T > 15$ GeV and an absolute value of pseudorapidity $|\eta| < 2.5$ because of the geometry of the inner tracker.

In Fig. 14, we show the distribution of the number of jets per event after the preselection, for the two samples, WH and dijet. On average, there are 3 jets per event in the WH sample and 2 in the dijet one. In Fig. 15 is shown the distribution of the transverse momentum p_T and the pseudorapidity η. Jets in the WH sample are a little more dispersed in the p_T range shown, while in the dijet sample they are more concentrated at low *pt*; the mean (RMS) of the distribution for the first one is ~ 39 (34) while for the second one it is ~ 23 (16).

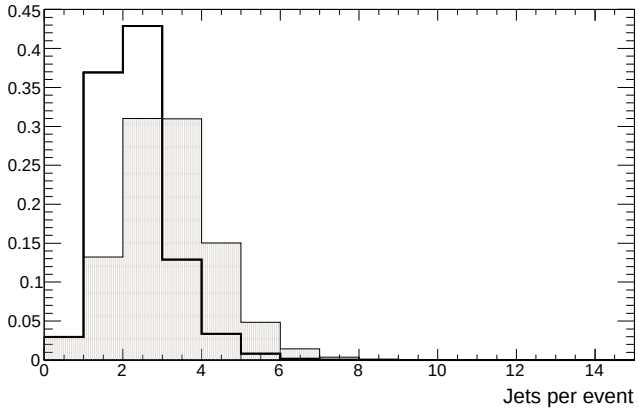

Figure 14: Distribution of the number of jets per event, for WH (hatched histograms) and dijet (solid line) samples.

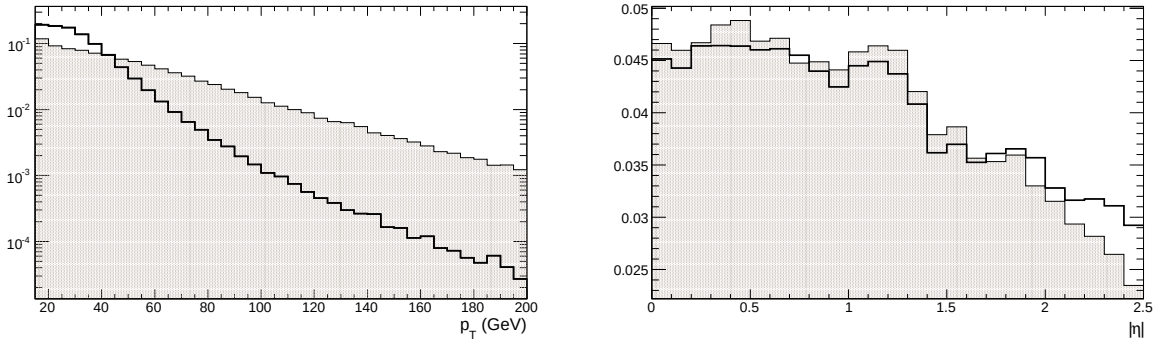

Figure 15: Distribution of transverse momentum p_T (left) and pseudorapidity η (right) for WH (hatched histograms) and dijet (solid line) samples.

5.2.2 Jets reconstruction performance

To measure the performance of the jets reconstruction, it has been used jets built from stable particles produced by the Monte Carlo generator. Stables particles are those that reach the detector without decaying first, which in ATLAS requires a laboratory frame lifetime of about ten picoseconds. These jets are referred to as *truth jets*.

In Fig. 16, it is shown the reconstruction efficiency ε_{rec} as a function of the transverse momentum $p_{T\,truth}$ and pseudorapidity η_{truth} of the *truth jets*. The reconstruction efficiency is defined as the ratio $N^{rec}_{truth}/N_{truth}$, where N_{truth} is the number of *truth jets* and N^{rec}_{truth} is the number

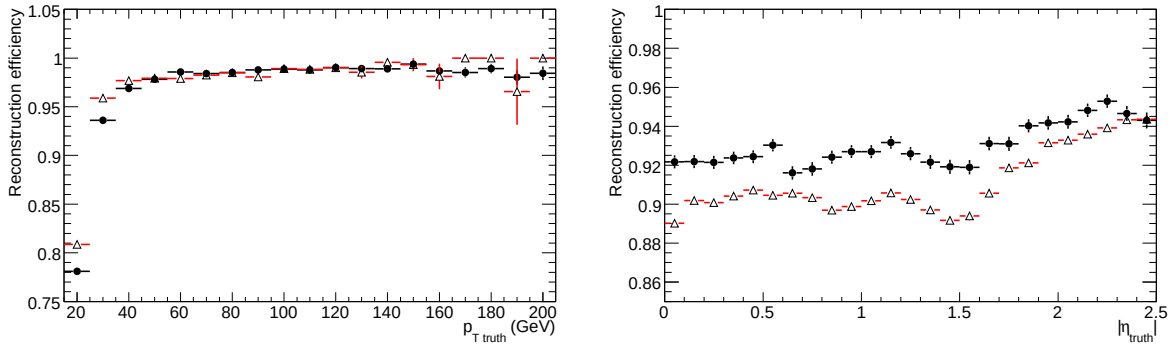

Figure 16: Jets reconstruction efficiency as a function of the transverse momentum (left) and pseudorapidity (right) for WH (circle) and dijet (triangle) samples. The total jets reconstruction efficiency in the p_T and η ranges of interest is $\varepsilon_j \sim 93\%$ for WH and $\varepsilon_j \sim 91\%$ for dijet samples.

of these ones that are reconstructed. The association between the *truth jets* and the reconstructed ones is done through directional matching, requiring a distance between them $\Delta R < 0.2$.

It can be seen that the jet reconstruction, for both samples, is highly efficient, except at very low transverse momentum. The total jets reconstruction efficiency in the p_T and η ranges of interest is $\varepsilon_j \sim 93\%$ for WH and $\varepsilon_j \sim 91\%$ for dijet samples.

In the same way the quality of the reconstructed variables, energy E_{rec} and pseudorapidity η_{rec}, has been briefly studied. Fig. 17 shows the ratio between the reconstructed and the *truth* energies E_{rec}/E_{truth} as a function of the *truth jets* energy (left) and their pseudorapidity (right). As well as the difference between η_{rec} and η_{truth} is shown in Fig. 18, as a function of *truth* pseudorapidity. In this figures one can see again that the jets reconstruction have a good performance. On average the reconstructed energy is above the 90% of the *truth energy*. Fig. 18 shows also the accurate of the polar coordinate η reconstruction.

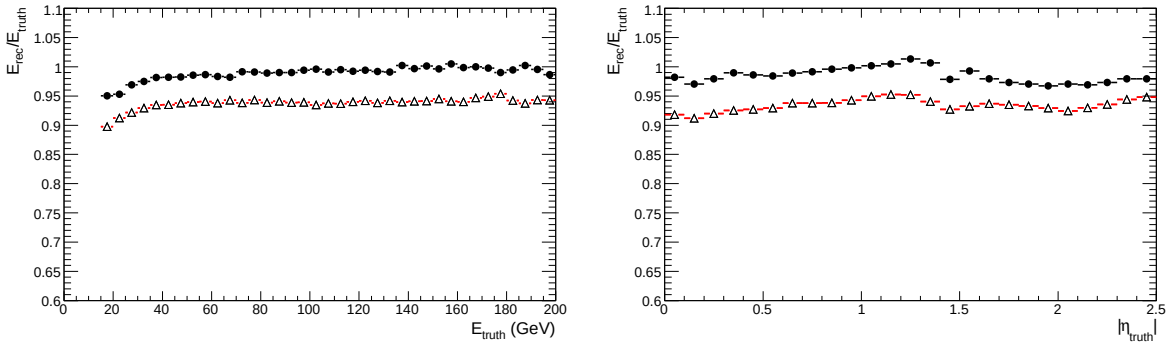

Figure 17: Reconstructed energy over *truth* energy versus *truth* energy (left) and the absolute value of the *truth* pseudorapidity (right) for WH (circle) and dijet (triangles) samples.

Purely electromagnetic jets

The jets reconstruction algorithm also reconstructs some electromagnetic particles, such as electrons and photons, as jet objects. In order to reject these fakes jets, we have plotted the distribution of the distance ΔR between a given jet and the closest reconstructed electromagnetic particles, versus the ratio between the particle energy and the jet energy. This distribution is shown in Fig. 19, considering electrons in the left side and photons in the right one. A jet is labeled as electromagnetic and rejected if $\Delta R < 0.2$, and the energy ratio is above 0.8. This labeling of electromagnetic jets was used by Cimmarusti for analysis of top decays [7].

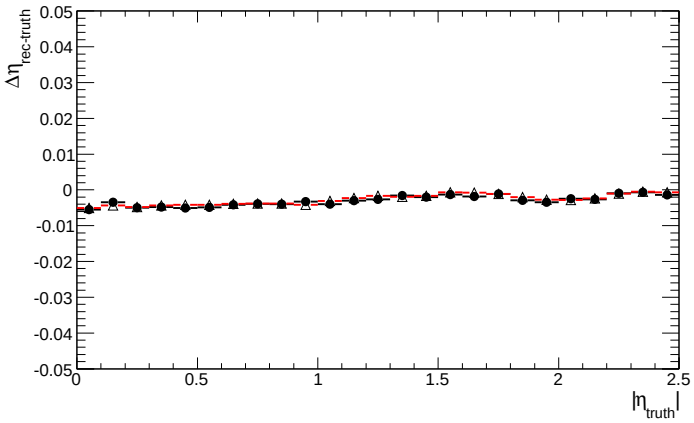

Figure 18: Difference between η_{rec} and η_{truth} versus the absolute value of η_{truth} for WH (circle) and dijet (triangles) samples.

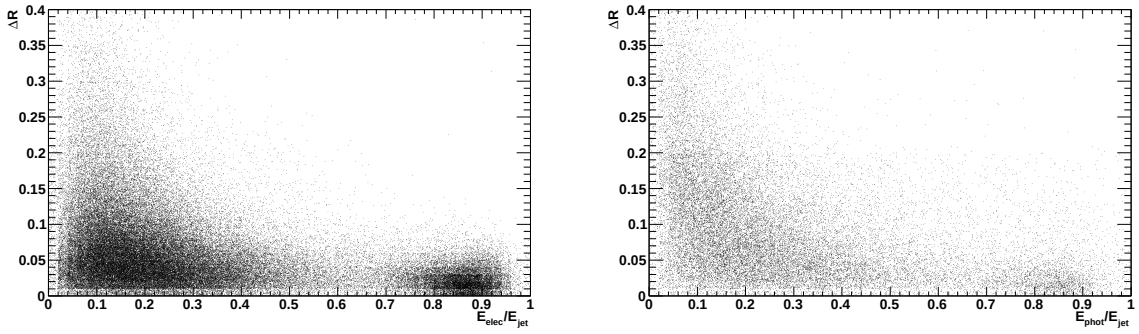

Figure 19: Distribution of the distance ΔR between a given jet and the closest reconstructed electromagnetic particles versus the ratio between the particle energy and the jet energy; for electrons (left) and photons (right), for the WH sample.

5.3 Jets signal and background

5.3.1 Jets labelling

In the ATLAS experiment, jets can be separated into four categories, jets originated from *b*-quarks, *c*-quarks, τ-lepton, and the light jets. The light jets are those originated from *u*, *d* and *s*-quarks or gluons. The Monte Carlo truth is used to label jets according to this classification, looking for heavy quarks (b and c) and τ leptons with $p_T > 5$ GeV within a cone $\Delta R < 0.3$ [15]. If there are more than one of them within the cone, the jet is considered to be a *b*-jet if the cone contains at least one *b*-quark. It is called a *c*-jet if there is at least one *c*-quark in the cone and no *b*-quark. In the same way for jets with τ-lepton and no *b* nor *c*-quark. And the remaining are considered as light jets.

5.3.2 Jets signal and background characteristics

In this analysis emphasis is done in the distinction of *b* jets from light jets. In Fig. 20, it can be compared the distributions of the transverse momentum p_T, energy E and pseudorapidity η for jets labelled as *b* and light jets, for both samples.

In general, one can see that the distribution of these variables are similar for *b*-jets and light jets in both samples; nevertheless, a higher concentration of light jets is seen at very low energy and p_T. Table 2 resumes the averages values for the energy and p_T distributions.

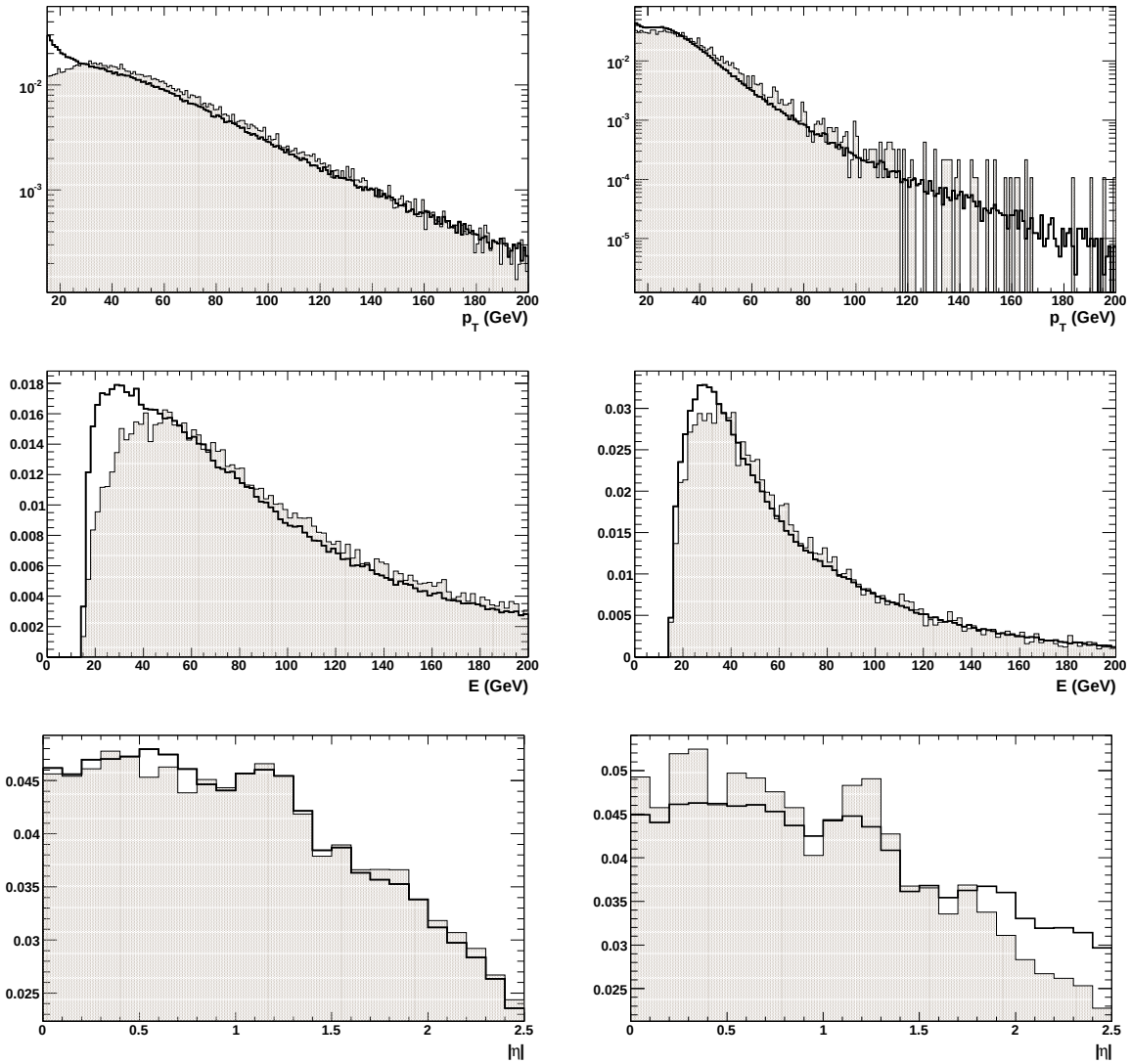

Figure 20: Distribution of the transverse momentum (top), energy (middle) and pseudorapidity (bottom) of reconstructed jets for WH (left) and dijet (right) samples. Distributions are shown for *b*-jets (hatched histograms) and light-jets (solid lines). Only jets with $p_T > 15$ GeV and $|\eta| < 2.5$ are considered.

Table 2: Average values of the labelled jets transverse momentum $< p_T^j >$ and energy $< E^j >$ (with corresponding RMS) for WH and dijets samples.

sample	Jet type	$< p_T^j > (RMS)$[GeV]	$< E^j > (RMS)$[GeV]
WH	b jets	57 (34)	86(46)
	light jets	53 (35)	79(47)
dijets	b jets	35 (17)	65(40)
	light jets	32(15)	64(41)

6 b-tagging studies

Many signatures of searches for a Standard Model Higgs boson, top physics and physics beyond the SM include presence of jets originating from b quarks; therefore an efficient identification of such jets is required. We have studied a method to tag b-jets by looking for the presence of an electron coming from the semileptonic decay of the B-hadrons. A description of the method and the results found are presented after an overview of the b-tagging methods used in ATLAS.

6.1 Overview of b-tagging algorithms

The most powerful way to tag b-jets is to take advantage of the relatively long lifetime of hadrons containing a b quark, of the order of 1.5 ps ($c\tau \approx 450\mu m$). Then, these hadrons have a significant flight path length before decaying and the tracks products of the decay will tend to have large transverse d_0 and longitudinal z_0 impact parameters. It can be used to distinguish between tracks from b quarks and tracks coming from lighter hadrons. The official ATLAS tagging algorithm use this fact to identify b-jets. Additionally, this algorithm reconstructs a secondary vertex with tracks preselected according to these parameters and calculates a likelihood function using the vertexed tracks invariant mass, the energy of these tracks and others [15, 16]. Typically, a b-tagging efficiency of about 60% is achieved with a rejection of light jets above one hundred.

On the other hand, the semileptonic decays of b-hadrons provide a clean signature used as well to identify b-jets. This is the so-called soft lepton tagging (the lepton being soft compared to high-p_T leptons from W or Z decays). The soft leptons tagging is not competitive with the lifetime tag because it is limited by the branching ratios of the decay to leptons (around 21% of B-mesons decays to leptons, per lepton family, including cascade decays). Using muons, a b-tagging efficiency $\varepsilon_b \sim 10\%$ have been achieved for a light jets rejection $R_{\text{light jet}} \sim 300$ [15]; while using electrons, that have a harder background, $\varepsilon_b \sim 7\%$ have been achieved with $R_{\text{light jet}} \sim 100$ [5].

Nevertheless, the soft lepton tagger has only small correlation with the lifetime tagger, which is very important for checking and cross-calibrating the b-tagging performance with real data.

6.2 Soft electron b-tagging algorithm

6.2.1 Electrons in jets reconstruction and identification

In our b-tagging analysis, the signal electrons come mainly from direct ($b \to e$) and cascade ($b \to c \to e$) semi-leptonic decays of B-hadrons and with less statistics from decays such as $b \to \tau \to e$ and $b \to (J/\psi, \psi') \to e^+e^-$ [2)]. The background electrons arise from π^0 Dalitz decays, γ-conversions occurring in the inner detector and decays of light hadrons. As well, pions misidentified as electrons represent background in the analysis. All these make the rejection of light jets difficult.

The standard electron reconstruction procedure (see section 4.3) is based on calorimeter clusters, with a subsequent association to tracks. While this method is efficient for high-energy

[2)] The corresponding branching ratios [17] are $Br(b \to \ell^-) = (10.71 \pm 0.22)\%$, $Br(b \to c \to \ell^+) = (8.01 \pm 0.18)\%$, $Br(b \to \bar{c} \to \ell^-) = (1.62^{+0.44}_{-0.36})\%$, $Br(b \to \tau \to e) = (0.419 \pm 0.055)\%$ and $Br(b \to (J/\psi, \psi') \to e^+e^-) = (0.072 \pm 0.006)\%$.

isolated electrons, such as those arising from W or Z decays, it is not effective for electrons inside hadronic jets, such as those from semi-leptonic decays, since their showers tend to overlap with the ones from hadrons in the collimated jets. An alternative procedure is used for soft electrons b-tagging [5] (the same is used for J/ψ decays analysis [4]). It takes full advantage of the tracking capabilities of the inner detector as well as the granularity of the electromagnetic calorimeter. The method relies on the extrapolation of reconstructed charged particle trajectories into the electromagnetic calorimeter.

In same way as for isolated electrons (section 4.3), the electrons identification algorithm incorporate variables that describe the shower shapes, quality of the match between the track and its corresponding cluster, and information from the transition radiation tracker. Additionally, for the identification is also used the electron momentum in the plane orthogonal to the jet axis (p_T^{rel}). It is because leptons from the semileptonic decays of heavy quarks are expected to have larger angles with respect to axes of jets than the ones from light quarks. In Fig. 21 is shown the distribution of p_T^{rel} for signals electrons and for pions in light jets.

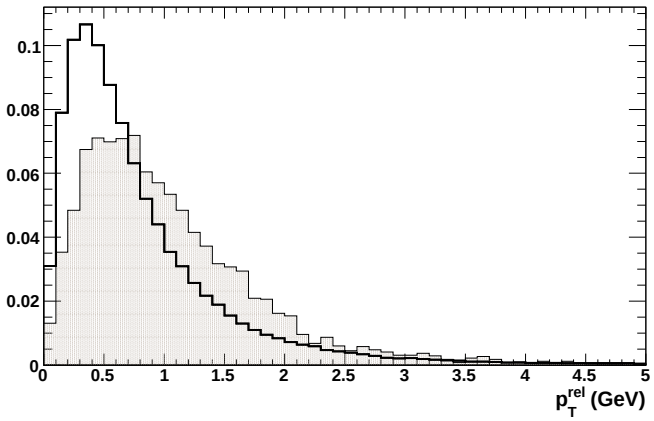

Figure 21: Distribution of the signal electrons momentum in the plane orthogonal to b-jets (hatched histogram) and the same variable for pions in light jets (solid line); for the WH sample.

6.2.2 b-tagging procedure

All the variables with capabilities to distinguish between signal electrons and non-electrons tracks or electrons from γ-conversions and Dalitz decays are combined in a discriminating function. It is, in fact, the likelihood for an objet reconstructed as electron to be a signal electron or background. The electron with the highest value of this function is chosen for each jet. Then the discriminating function of the electron will represent the likelihood for the jet to be a b-jet. In Fig. 22 is shown the distribution of this discriminating function D_{jet} for b-jets and light jets. For a given threshold D_{jet}^{thr}, a jet with $D_{jet} > D_{jet}^{thr}$ is tagged as a b-jet.

6.2.3 b-tagging performance

The b-tagging efficiency is defined as $\varepsilon_b = N_b^t/N_b$ where N_b^t is the number of tagged b-jets and N_b is the total number of b-jets. This definition includes the semi-leptonic branching ratios as well as the electron reconstruction and identification efficiencies. The light jet rejection factor

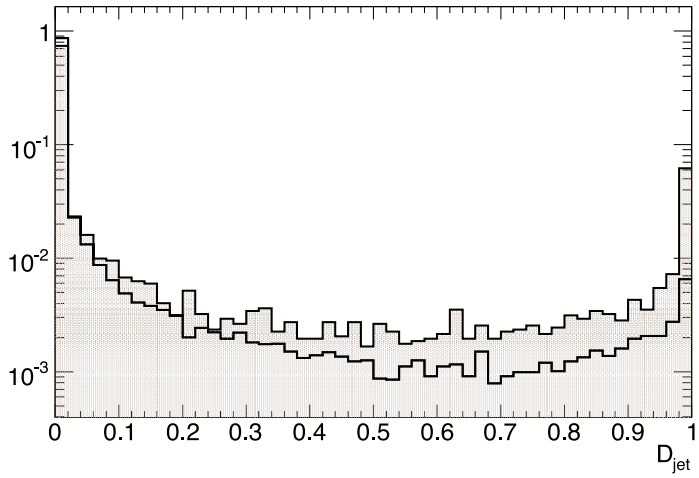

Figure 22: Distribution of the discriminating function for *b*-jets (hatched histogram) and light jets (solid line); for the WH sample.

is calculated as $R_{\text{light jet}} = N_{\text{light}}/N^t_{\text{light}}$, where N_{light} is the number of light jets and N^t_{light} is the number of light jets tagged by mistake as *b*-jets.

Fig. 23 shows the rejection of light jets as a function of the *b*-tagging efficiency ε_b, for the *WH* sample in the left side and for the *dijet* sample in the right side. The results obtained with both samples are very similar. In both case a rejection of light jets of about 140 is achieved with a b-tagging efficiency of about 7%.

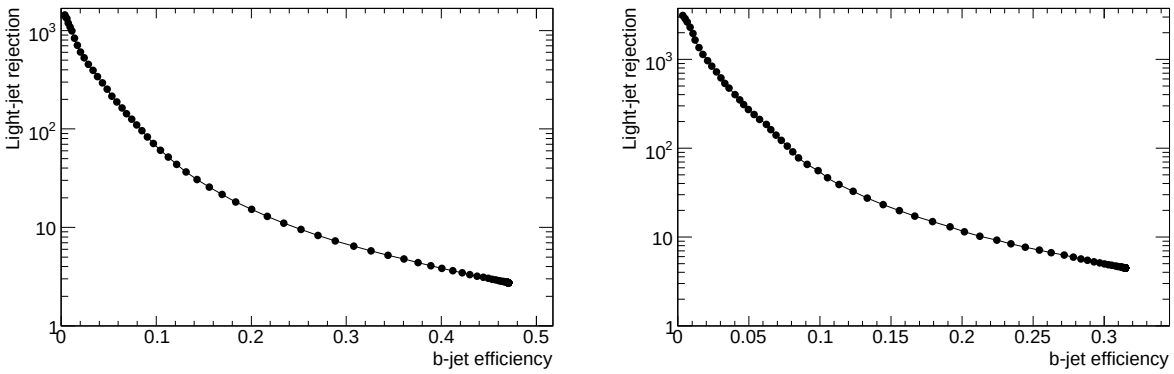

Figure 23: Rejection of light jets versus *b*-tagging efficiency ε_b, for *WH* events in the left side and for *dijet* events in the right side.

In a previous study [5], the performance of this algorithm had been measured for *WH* samples obtained with the previous version of the ATLAS simulations and reconstruction software. The result obtained in that analysis was a rejection slightly lower (~ 110) for the same value of b-tagging efficiency. In the case of the dijet sample, this is first time that this measurement have been done.

6.3 Calibration of *b*-tagging using dijets

In the analysis of many physics processes where *b*-tagging is required, in order to estimate the background after applying *b*-tagging algorithms, it is necessary to know the tagging efficiency for *b*-jets, for light jets (mistagging rate) and for *c*-jets (*c*-tagging efficiency) with high accuracy. Until now the Monte Carlo simulations have been used for the estimation of possible background in this processes, nevertheless, It is clear that for those analysis there will be a need to calculate these efficiencies for real data.

In ATLAS the technique System 8 will be used for this purpose [16]. In this method two data samples with different *b*-fractions are considered, to which one applies two different *b*-tagging algorithms, the lifetime tagger (described in section 6.1) and a soft lepton tagger (here is where our tagger plays an important role). For each sample four numbers are obtained, the number of jets before tagging, the number of jets tagged by the lifetime tagger, the number of jets tagged by the soft lepton tagger and the number of jets tagged by both. Then, a system of eight equations with eight unknowns can be wrote (hence the name of the method), where some of the unknowns are the quantities that one is looking for, the efficiencies of the tagging algorithms.

On the other hand, we already know that at the beginning at the LHC, there will be a large amount a of dijets events, therefore, samples of these events could be used to study *b*-tagging with real data and for the calibration of these algorithms. At the end of the internship some preliminaries studies were done, looking for a way to enrich a *dijets* sample with events containing *b*-jets with the *b* decaying semi-electronically. Since there is not jet plus electron trigger available, only electron triggers have been studied. It has been found that a trigger menu that requires at least one electron with $E_T > 10$ GeV could be used with some promising results. From a sample that contain originally $\sim 0.6\%$ of dijets events with $b \to e\nu X$, using this trigger menu, one can obtain a sample with $\sim 13\%$ of events with the desired topology. It is reducing the samples to only 0.5% of its initial size. Nevertheless, this trigger menu will be available only at the begging and not in forwards periods with high luminosity; therefore, others ways have to be studied.

7 Conclusion

During this HELEN internship, a new framework to do soft electron b-tagging has been developed. It works with the ATLAS analysis codes using samples of data in the AOD format (Analysis Object Data), and provides ROOT tuples to make histograms.

The performance of the soft electron b-tagging algorithm has been measured, using the developed framework, for the WH and dijets samples. For both samples the b-tagging efficiency is about 7% with a rejection of light jets of ~ 140. This efficiency includes the branching ratio for the semileptonic decay of the *B* hadrons (about 20% for each lepton family) and the electron reconstruction efficiency.

Some preliminaries studies have been done in order to enrich dijets samples in events containing b-jets with $b \to e\nu X$. Using a trigger menu that requires at least one electrons with $E_T > 10$ GeV, a sample can be enriched to have 13% of its events with the desired topology, being this fraction of about 0.6% before the application of the trigger. This trigger menu will be available during the period of low luminosity at the LHC.

8 Acknowledgements

I acknowledge to the HELEN project for giving me the opportunity of doing this internship and to José Ocariz and Luis Núñez for their inconditional support. I am equally grateful to the LPNHE laboratory for reciving me in its framework group and, specially, to Frederic Dereu for his valuable steering and teaching.

References

[1] HELEN web page: http://www.roma1.infn.it/exp/helen/.

[2] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST, 2008.

[3] Derue F., Serfon C., *Electron-jet separation with DC1 data*, 2005, ATL-PHYS-PUB-2005-0016.

[4] ATLAS Collaboration, *Reconstruction of $J/\psi \to e^+e^-$ and $\Upsilon \to e^+e^-$ decays*, 2008, ATLAS CSC note.

[5] ATLAS Colaboration, *Soft electron b-tagging with CSC data*, 2008, ATLAS CSC note.

[6] Fundamental Particles and Interactions Poster. Particle Adventure web page: http://www.particleadventure.org/.

[7] Cimmarusti, A., *Analysis of top anti-top decays in the lepton plus jets channel at the ATLAS experiment*, 2006, HELEN fellowship.

[8] Lyndon Evans and Philip Bryant, *LHC Machine*, JINST, 2008.

[9] ATLAS web page: http://atlas.ch/.

[10] Derue F., *Reconstruction et identification des électrons dans l'expérience Atlas. Participation á la mise en place d'un Tier 2 de la grille de calcul.*, 2008.

[11] The ATLAS Twiki web page: https://twiki.cern.ch/twiki/bin/view/Atlas/WebHome.

[12] Worldwide LHC Computing Grid web page: http://lcg.web.cern.ch/LCG/mou.htm.

[13] C. Amsler, *et al.*, PDG Particle Physics Booklet, 2008.

[14] Cacciari, M. and Salam, G.P., Phys. Lett. **B641** (2006) 57–61.

[15] ATLAS Collaboration, *b-tagging performance*, 2008, ATLAS CSC note BT-0.

[16] ATLAS Colaboration, *Calibrating the ATLAS b-tagging Algorithms Using Dijet Events*, 2008, ATLAS CSC note.

[17] W.-M.Yao et al. (Particle Data Group), J. Phys. G 33, 1 (2006) and 2007 partial update for the 2008 edition.