

Double  $B$ -hadron Jet Tagging and  
Identification of Gluon to  $b\bar{b}$  jets with the  
ATLAS Detector

Lic. María Laura González Silva

Tesis Doctoral en Ciencias Físicas  
Facultad de Ciencias Exactas y Naturales  
Universidad de Buenos Aires

Noviembre 2012



**UNIVERSIDAD DE BUENOS AIRES**

Facultad de Ciencias Exactas y Naturales

Departamento de Física

**Double  $B$ -hadron Jet Tagging and Identification of  
Gluon to  $b\bar{b}$  jets with the ATLAS Detector**

Trabajo de Tesis para optar por el título de  
Doctor de la Universidad de Buenos Aires en el área Ciencias Físicas

por      **María Laura González Silva**

Director de Tesis: Dr. Ricardo Piegaia

Consejero de estudios: Dr. Daniel Deflorian

Lugar de Trabajo: Departamento de Física (CONICET-UBA)

Buenos Aires, 2012

## **AGRADECIMIENTOS**

A Ricardo, a mi familia...

## Abstract

El detector ATLAS, uno de los cuatro experimentos del acelerador LHC actualmente en construcción en el CERN, tiene como propósito principal el descubrimiento del bosón de Higgs, la partícula involucrada en el mecanismo de ruptura de la simetría electrodébil y el origen de la masa, así como la búsqueda de nueva física más allá Modelo Estándar hasta la escala de 1 TeV. En el colisionador chocarán haces de protones con una energía de centro de masa de 14 TeV y una frecuencia de interacción de  $10^9$  Hz.

This thesis describes a method that allows the identification of double  $B$ -hadron jets originating from gluon-splitting. The technique exploits the kinematic differences between the so called “merged” jets and single  $B$ -hadron jets using track-based jet shape and jet substructure variables combined in a multivariate likelihood analysis. The ability to reject  $b$ -jets from gluon splitting is important to reduce and to improve the estimation of the  $b$ -tag background in Standard Model analyses and in new physics searches involving  $b$ -jets in the final state. In the simulation, the algorithm rejects 95% (50%) of merged  $B$ -hadron jets while retaining 50% (90%) of the tagged  $b$ -jets, although the exact values depend on the jet  $p_T$ .

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Identification of $b$ -jets from gluon splitting . . . . .	4
1.2	Physics Motivation . . . . .	6
<b>2</b>	<b>The theory of the strong interactions</b>	<b>13</b>
2.1	The parton model of QCD . . . . .	13
2.2	Jet physics . . . . .	13
2.2.1	Monte Carlo simulations . . . . .	13
2.2.2	Jet algorithms . . . . .	13
2.2.3	Jet substructure . . . . .	13
<b>3</b>	<b>The ATLAS detector at the LHC</b>	<b>14</b>
3.1	The Large Hadron Collider . . . . .	14
3.1.1	Luminosity and pile-up . . . . .	18
3.2	The ATLAS Detector . . . . .	21
3.2.1	Inner tracking system . . . . .	23
3.2.2	The Calorimeter System . . . . .	27
3.2.3	The Muon System . . . . .	32
3.2.4	Trigger and Data Adquisition . . . . .	35
3.2.5	El HLT . . . . .	38

3.2.6	Data quality . . . . .	43
3.2.7	Simulation of particle interactions in the ATLAS Detector . . . . .	43
<b>4</b>	<b>Jet reconstruction and <math>b</math>-Tagging</b>	<b>44</b>
4.1	Jets reconstruction and calibration . . . . .	44
4.2	$b$ -jet Tagging . . . . .	45
<b>5</b>	<b>Double <math>B</math>-hadron jet identification</b>	<b>46</b>
5.1	Data and Monte Carlo samples . . . . .	46
5.2	Event and object selection . . . . .	48
5.3	Preliminary studies in standalone PYTHIA generator . . . . .	50
5.4	Kinematic differences between single and double $B$ -hadron jets	53
5.4.1	Further studies using “ghost-association” and bigger cone jets . . . . .	64
5.5	Validation of the jet variables in data . . . . .	66
<b>6</b>	<b>Multivariate Analysis</b>	<b>70</b>
6.1	The multivariate classifiers . . . . .	70
6.2	The input variables . . . . .	74
6.3	$g \rightarrow b\bar{b}$ likelihood training and performance . . . . .	74
6.4	Systematic uncertainties . . . . .	77
6.5	Isolation studies . . . . .	82
6.6	Other Monte Carlo generators . . . . .	83
<b>7</b>	<b>??Fraction of gluon-splitting jets in data??</b>	<b>86</b>
7.1	Template fits?? . . . . .	86
<b>8</b>	<b>Conclusions</b>	<b>87</b>

%

# Chapter 1

## Introduction

### 1.1 Identification of $b$ -jets from gluon splitting

The ability to identify jets containing  $B$ -hadrons is important for the high- $p_T$  physics program of the ATLAS experiment. Two robust  $b$ -tagging algorithms taking advantage of  $B$ -hadron pair via gluon splitting, as depicted in Fig. 1.1. We will henceforth call “merged”  $b$ -jets  $b\bar{b}$  jets the  $b$ -tagged jets containing two  $B$ -hadrons. The ability to single out  $b$ -tagged jets from gluon splitting has several applications in different lines of analysis: measurement of QCD beauty production,  $t\bar{t}$  and single top production, reduction of background in searches with  $b$ -quarks in the final state, and the study of substructure in fat jets, where  $g \rightarrow b\bar{b}$  jets compete with boosted  $Z \rightarrow b\bar{b}$  and  $H \rightarrow b\bar{b}$  jets.

There are two possible strategies to attempt the identification of  $b$ -jets containing two  $B$ -hadrons. One of them relies on the direct reconstruction of the two  $B$ -decay secondary vertices [4]. This has the further advantage of allowing the measurement of the angular separation between the  $B$ -hadrons,

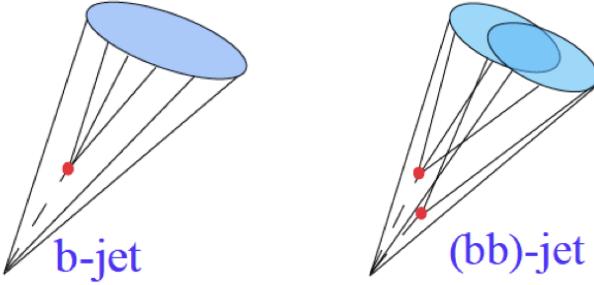


Figure 1.1:  $b$ -tagging algorithms select jets originating both from the fragmentation of a single  $b$  quark (“single”  $b$ -jets, left image) or from the splitting of a gluon into a pair of close-by  $b\bar{b}$  quarks (“merged”  $b$ -jets, right image).

but suffers from the low efficiency of a double  $b$ -tag requirement plus additional reconstruction inefficiencies at small angular separation between the two  $B$ s. In this paper we develop an alternative method that does not rely on explicitly finding vertices, but exploits the substructure differences between single and merged  $b$ -jets, combining them in a multivariate analysis.

The note is organized as follows. In section 1.2 we review the physics cases where this tool finds natural applications. Sections 5.1 and 4.1 review the Data and Monte Carlo samples and their reconstruction and section 5.2, the criteria for selection of events, jets and tracks. The kinematic variables that differentiate between single and merged  $b$ -jets are discussed in section ?? and the validation of their MC distributions with QCD data in section ???. The construction of the multi-variate discriminator is presented in section ?? and the discussion of the systematic uncertainties in section ???. Section 6.6 investigates the performance of the tagger with other Monte Carlo generators and, finally, section ?? summarizes the results and discusses future improvements and new ideas.

## 1.2 Physics Motivation

Within the Standard Model (SM) a range of production channels exist for heavy-quark jets, pure QCD production or in association with heavy bosons ( $W, Z, H$ ). Furthermore,  $b$ -quarks enter in many collider searches, notably because they are produced in the decays of various SM particles, top quarks and the Higgs boson (if light), and of numerous particles appearing in proposed extensions of the SM. The ability to distinguish genuine  $b$ -quark jets from those produced via gluon splitting is thus of wide application. Here we briefly discuss three cases, the measurement of QCD  $b$  quark production, the reduction of background in SM and BSM analyses with  $b$  quarks in the final state, and studies of jet substructure.

### *The measurement of the inclusive $b$ -jet spectrum*

The simplest and most fundamental measurement of heavy-quark jet production is the inclusive heavy-quark jet spectrum, which is dominated by pure QCD contributions. Studies of QCD bottom jets production are of intrinsic interest because of the correspondence between parton level production and the observed hadron level, and their potential to provide information on the  $b$ -quark parton distribution function, the only component of the proton structure thought to be generated entirely perturbatively from the DGLAP evolution of the other flavours. The theoretical calculation of the inclusive  $b$ -jet spectrum presents the striking feature of having rather important uncertainties ( $\sim 50\%$ ), considerably larger than the corresponding ones for the normal (light) jet inclusive spectrum ( $\sim 10\text{-}20\%$ ), see for example [5].

The origin of these uncertainties are reviewed in a recent paper by Banfi, Salam and Zanderighi [6], from which we have taken Fig. 1.2. Its top panel

shows the  $K$ -factor, the ratio of the next-to-leading order (NLO) to the leading order (LO) cross section, obtained with MCFM for the LHC design energy ( $pp, \sqrt{s} = 14$  TeV). The fact that NLO terms are considerably larger than the LO ones indicates that the perturbative series is very poorly convergent, and implies that the NLO result cannot be an accurate approximation to the full result. It is for this reason that the scale dependence (middle panels) is large. The poor convergence of the perturbative series is related to the different channels for heavy quark production. At leading order only the so-called flavour creation channel (FCR) is present,  $\ell\ell \rightarrow b$ , where  $\ell$  is a generic light parton (quark or gluon), see fig. 1.3. At NLO, two new channels open up, often referred to as flavour excitation (FEX) and gluon splitting (GSP). In the former, a gluon from one of the incoming hadrons splits collinearly into a  $b\bar{b}$ -pair and one of those  $b$ -quarks enters the hard  $b\ell \rightarrow b\ell$  scattering. In the gluon splitting process, the hard scattering LO diagram is of the form  $\ell\ell \rightarrow \ell\ell$ , and one of the final-state light partons (at NLO always a gluon) splits collinearly into a  $b\bar{b}$ -pair that the clustering algorithm can classify within the same jet. A jet containing both  $b$  and  $\bar{b}$  is considered to be just a  $b$ -jet in standard definitions. The various channels can be conveniently separated with a parton shower Monte Carlo generator such as Herwig [8], where one can determine the underlying hard channel from the hard process in the event record. Their relative contributions to the total  $b$ -jet spectrum are shown in the bottom panel of fig. 1.2. One sees that the supposedly LO channel (FCR) is nearly always smaller than the two channels that enter only at NLO (FEX and GSP).

The largest residual uncertainties are associated with the channel with the most logarithms, gluon splitting. This channel however does not even correspond to one's physical idea of a  $b$ -jet, the one induced by a hard  $b$ -quark,

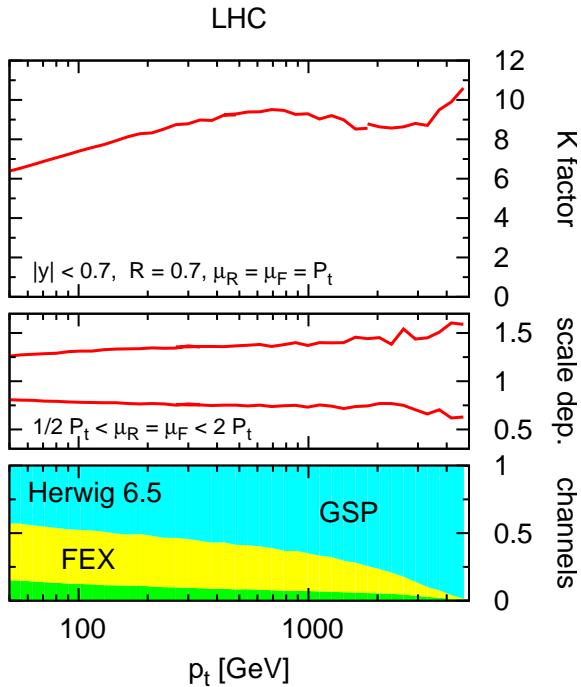


Figure 1.2: Top:  $K$ -factor for inclusive  $b$ -jet spectrum taken from [6], clustering particles into jets using the  $k_t$  jet-algorithm [7] with  $R=0.7$ , and selecting jets in the central rapidity region ( $|y| < 0.7$ ). Middle: scale dependence obtained by simultaneously varying the renormalisation and factorisation scales by a factor two around  $p_T$ , the transverse momentum of the hardest jet in the event. Bottom: breakdown of the Herwig [8] inclusive  $b$ -jet spectrum into the three major underlying channels, flavor creation (FCR) flavor excitation (FEX) and gluon splitting (GSP).

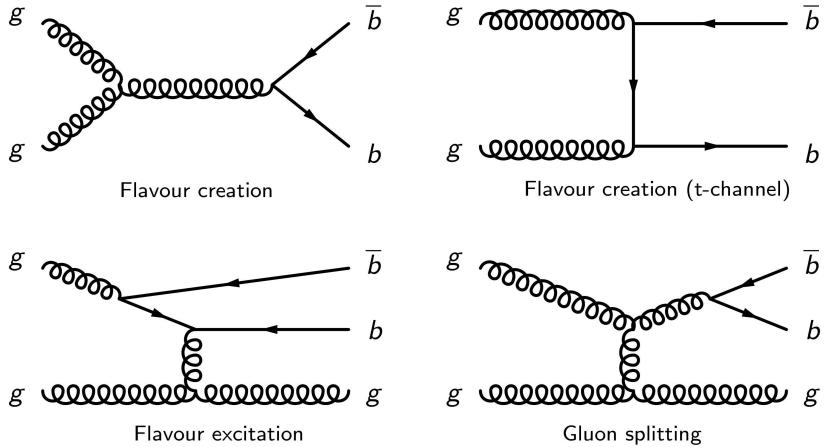


Figure 1.3: Typical Feynman diagrams for the 3 processes contributing up to next-to-leading order to QCD bottom production: (a-b) Flavor creation (FCR), (c) Flavor excitation (FEX), (d) Gluon splitting (GSP).

and it seems somehow unnatural to include it at all as part of one's  $b$ -jet spectrum. Reference [6] thus proposes a new approach to improving the accuracy of the prediction of the  $b$ -jet spectrum, where  $b$ -jets definition maintains the correspondence between partonic flavour and jet flavour. Specifically, a jet containing equal number of  $b$ -quarks and  $b$ -antiquarks is considered to be a light jet, so that jets identified as gluon splitting are removed from the  $b$ -jet spectrum.

#### *Rejection of background in SM analyses and beyond-SM searches*

Efficient tagging of merged  $b$ -jets from gluon splitting can provide an important handle to understand, estimate and/or reject  $b$ -tag backgrounds to Standard Model and new physics searches at the LHC.

Standard Model physics analyses that rely on the presence of  $b$  quarks in the final state such as top quark physics, either in the  $t\bar{t}$  or the single top

channels, and associated Higgs production:  $WH \rightarrow \ell\nu b\bar{b}$  and  $ZH \rightarrow \nu\nu b\bar{b}$ , suffer from reducible backgrounds from QCD (that can produce  $b$ -jets from gluon splitting) and, most importantly, from the irreducible background due to  $W$  bosons produced in association with  $b$  quarks. Fig. 1.4 shows the two leading order processes that give rise to  $W$  bosons with at least one  $b$ -jet. In the first process, which can be thought of as a higher order correction to  $W +$  jets production, the  $b$  quark pair is produced at small angles by gluon splitting and can often be reconstructed as a merged jet. NLO calculations

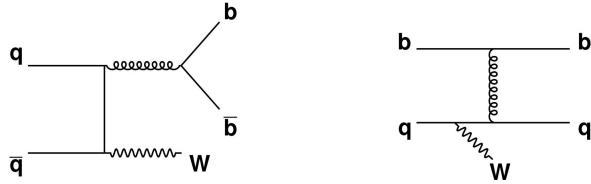


Figure 1.4: Leading order Feynman diagrams for  $W$  production in association with  $b$  quarks.

of the production of  $W$  bosons and two jets with at least one  $b$  quark at the LHC for jet  $p_T > 25$  GeV, and  $|\eta| < 2.5$  [9] indicate that the cross section for  $W(b\bar{b})j$  is almost a factor of two higher than  $Wb\bar{b}$ , and about a third of  $Wbj$ , where  $W(b\bar{b})j$  denotes the case in which the two  $b$  quarks are merged into the same jet.

New physics searches with  $b$  quarks in the final state can also greatly benefit from rejection of QCD and  $W + b$  backgrounds with  $b$ -jets arising from gluon splitting. For example consider the search for supersymmetry in the  $+ b$ -jets channel [10]. Within the framework of a generic  $R$ -parity conserving minimal supersymmetric extension of the SM The coloured superpartners of quarks and gluons, the squarks and gluinos, are expected to be copiously produced via the strong interaction at the LHC. The partners

of the right-handed and left-handed quarks, qR and qL, can mix to form two mass eigenstates, and these mixing effects being proportional to the corresponding fermion masses, they are expected to become most important for the third generation to yield sbottom ( $b_1$ ) and stop ( $t_1$ ) mass eigenstates significantly lighter than other squarks. Both sbottom and stop chain decay to  $b$  quarks and the lightest supersymmetric particle, producing the expected signal of +  $b$ -jets.

### *Jet substructure and boosted objects*

At the LHC, many of the particles considered to be heavy at previous accelerators will be frequently produced with a transverse momentum greatly exceeding their rest mass. Good examples are the electro-weak gauge bosons  $W^\pm$  and  $Z^0$ , the top quark, the Higgs boson or bosons and possibly other new particles in the same mass range. These boosted objects, produced either because they recoil against other energetic objects or because they arise from decays of even heavier BSM particles, can form upon decay a highly collimated topology too close to be resolved by a jet algorithm. For these cases, sophisticated tools have been developed in the last years [11, 12] to analyse the substructure of the ensuing jet and reveal its heavy-particle origin.

The study of  $b\bar{b}$  jets from gluon splitting is an ideal testbed for studying jet substructure in data, as it provides a large supply of boosted, merged jets. Furthermore, understanding  $g \rightarrow b\bar{b}$  jets is important as they are themselves the background to boosted object searches, like  $Z \rightarrow b\bar{b}$  or  $H \rightarrow b\bar{b}$ . In particular, it has recently been suggested [13] that  $WH$  and  $ZH$  production become potential discovery and analysis channels by restricting ones attention to the  $\sim 5\%$  of events in which the vector and Higgs bosons have large

transverse momentum,  $p_{TH} > 200$  GeV. Understanding the much more common QCD events with merged  $b\bar{b}$  jets will be essential before attempting to measure these rare final states.

# Chapter 2

## The theory of the strong interactions

### 2.1 The parton model of QCD

### 2.2 Jet physics

#### 2.2.1 Monte Carlo simulations

#### 2.2.2 Jet algorithms

#### 2.2.3 Jet substructure

# Chapter 3

## The ATLAS detector at the LHC

### 3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [14] is a proton-proton ( $pp$ ) synchrotron located in the previous Large Electron Positron (LEP) collider tunnel at CERN Laboratory, just outside the city of Geneva (Switzerland), approximately 100 m underground. It is designed to collide bunches of up to  $\sim 10^{11}$  protons every 25 ns at a center-of-mass energy of 14 TeV (seven times the 2 TeV reached by the Tevatron accelerator at Fermilab Laboratory, in Chicago).

The experiments analyzing the collisions produced by the LHC are distributed around the 27 km ring at the various interaction points. The ATLAS experiment is located at Point 1, which is closest to the main CERN site. Point 5 houses the other general purpose detector, CMS. ALICE and LHCb experiments are located at Point 2 and Point 8, respectively. The former is designed to investigate heavy ion collisions; the latter, to investigate rare decays of b-mesons. The layout of these four experiments along the LHC

ring is shown in Fig. 3.1.

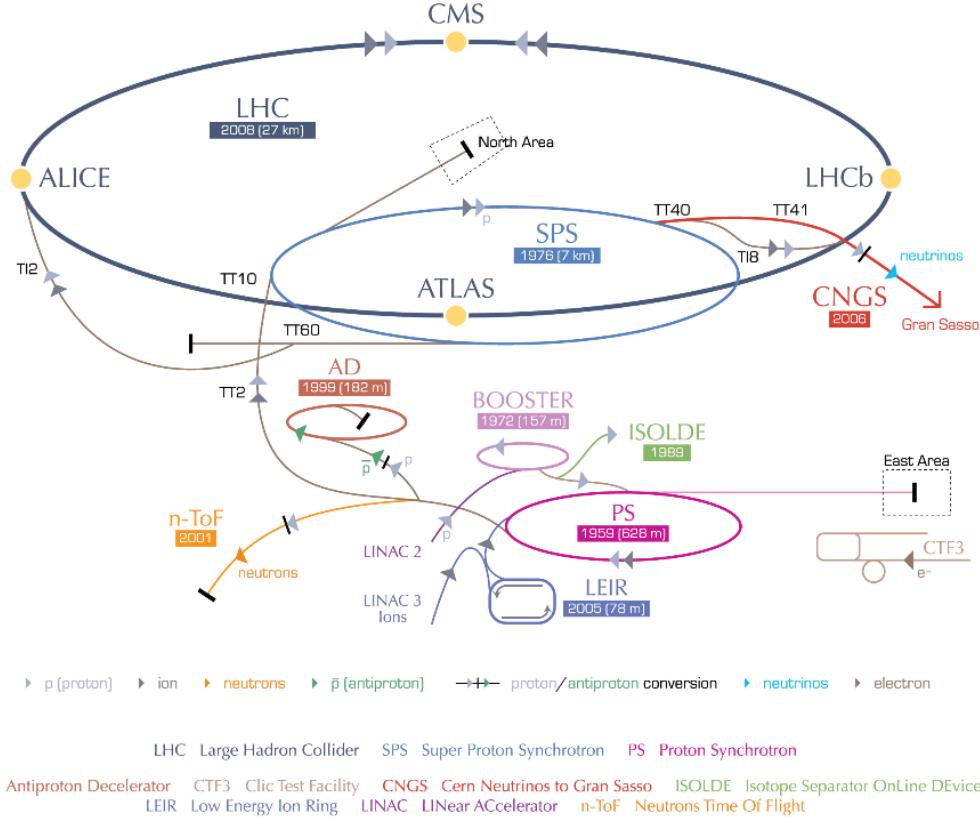


Figure 3.1: The CERN accelerator complex, showing the injection system, along with each components date of construction, and the placement of the four main experiments.

Proton beams are formed, before insertion into the main LHC ring, using a succession of smaller machines with increasingly higher energies, as shown in Fig. 3.1. The chain begins as protons are injected into the PS Booster (PSB) at an energy of 50 MeV from Linac2. The booster accelerates them to 1.4 GeV. The beam is then fed to the Proton Synchrotron (PS) where it is accelerated to 25 GeV. At desin strength, the bunch structure, known as a

bunch train, contains 72 bunches of protons upon entry to the Super Proton Synchrotron (SPS). The SPS accumulates up to four fills of 72 bunches from the PS and accelerates them to 450 GeV, with a bunch spacing of  $\sim 25$  ns. They are finally transferred to the LHC (both in a clockwise and an anticlockwise direction) where they are accelerated for 20 minutes to their nominal energy of 7 TeV. Beams will circulate for many hours inside the LHC beam pipes under normal operating conditions.

The bunch structure is a direct consequence of a radio frequency (RF) acceleration scheme used to attain the desired high proton beam energy. In RF acceleration, particles travel through a series of time-varying electrical fields and they can only be accelerated when the RF field has the correct orientation when particles pass through an accelerating cavity, which happens at well specified moments during an RF cycle. The result of a sequence of RF accelerations is several bunches of protons. It is important to note that when we speak about “beams” we refer to many bunches of protons separated by some uniform distance. Increasing the number of bunches is one of the ways to increase luminosity in a machine (more about luminosity in subsection 3.1.1). At desinged beam intensitiy, when the bunches cross, there will be a maximum of about 20 collisions.

A larged magnetic field is needed to guide and maintain the beam particles in their circular orbit. The needed field is achieved using superconducting electromagnets built from NbTi coils that operate in a superconducting state, efficiently conducting electricity without resistance or loss of energy. The currents through the coils produce magnetic fiedls perpendicular to the direction of motion of the protons that deflect the protons into their orbits. The whole magnetic system comprises 1232 dipole magnets of 15 m length which are used to bend the beams, and 392 quadrupole magnets, each 57 m

long, to focus the beams. At a peak beam energy of 7 TeV, the dipoles need to produce an 8.33 T magnetic field, requiring a current of  $\sim 12$  kA. In order to deliver the current densities and magnetic field required for 7 TeV proton beams, the magnets are kept at 1.9 K by circulating superfluid helium.

The first  $pp$  collisions produced by the LHC occurred on November 23 2009, at the SPS extraction energy of 450 GeV per beam. Very quickly after, on December 8, ATLAS and CMS detectors started recording data at energy of 2.36 TeV. By this time the LHC became the highest energy accelerator in the world. During this period, bunch intensities were limited by machine-protection considerations to  $1.5 \times 10^{10}$  protons.

In February 2010, the LHC was commissioned once more with 450 GeV beams, and a series of tests were performed to ensure that the magnet systems could operate safely at the currents necessary to control 3.5 TeV beams. This was followed by the very first collisions at 7 TeV center-of-mass energy on March 30. During the 2010 run the beam parameters were tuned (the beam widths squeezed and the number of protons per bunch and the number of bunches in each beam increased) in order to increase the beam intensity. In particular, as the intensity of the beams increased, the mean number of interactions per bunch crossing increased.

Finally, the data samples analysed in this thesis correspond to proton-proton collisions at  $\sqrt{s} = 7$  TeV delivered by the LHC and recorded by ATLAS between May and November 2011, with the LHC running with 50 ns bunch spacing. Table ?? summarizes the basic beam parameters expected for design energy and luminosity and the beam parameters as of May 2011. The LHC performance steadily improved during 2011. The average number of interactions per bunch crossing throughout the data-taking period considered rapidly increased approximately from  $\sim 3$  to 8 until (northern hemisphere)

summer 2011, with a global average for this period of  $\approx 6$ . Starting in August 2011 and lasting through the end of the proton run, this number ranged from approximately 5 to 17, with an average of about 12. This evolution is illustrated in Fig. 3.2, which shows the maximum mean number of collisions per beam crossing versus day in 2011.

Parameter	2011 runs	Design
Center-of-mass energy [TeV]	7	14
Instantaneous luminosity [ $\text{cm}^{-2}\text{s}^{-1}$ ]	$3.65 \times 10^{33}$ (year peak)	$10^{34}$
Bunches per beam	38 (May)	2808
Protons per bunch	$0.8 \times 10^{11}$ (May)	$1.5 \times 10^{11}$
Mean interactions per crossing	6 to 12 (year average)	23

Table 3.1: Summary of beam conditions during the 2011 7 TeV runs and those foreseen at design energy and luminosity.

### 3.1.1 Luminosity and pile-up

The rate of events produced by the colliding beams depends on the luminosity of the collisions, which is a measure of the number of events per second per unit cross section, typically measured in units  $\text{cm}^2\text{s}^{-1}$ . The number of events of a particular process, then, is given by the product of the integrated luminosity,  $\int dt L$ , and the cross section of the process,  $\sigma_{\text{event}}$ . The integrated luminosities are typically quoted in units of inverse picobarns,  $\text{pb}^{-1} = 10^{-36}\text{cm}^2$ . In order to measure processes with very little cross sections a very high luminosity is required.

The delivered luminosity can be written as [15]:

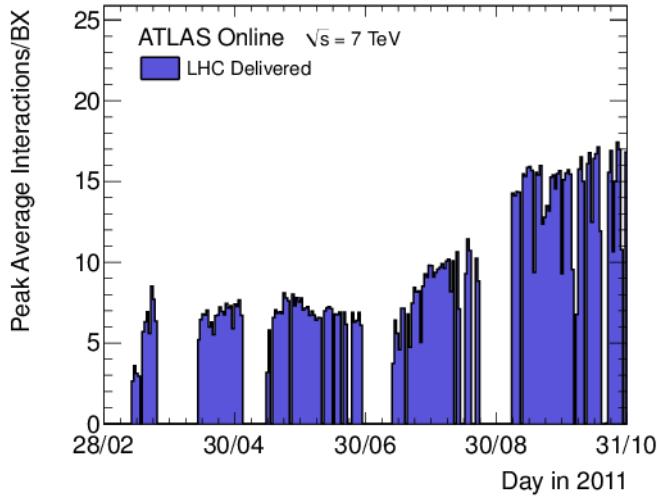


Figure 3.2: The maximum mean number of events per beam crossing versus day in 2011.

$$L = \frac{n_b f_r n_1 n_2}{2\pi \Sigma_x \Sigma_y} \quad (3.1)$$

where  $n_b$  is the number of colliding bunch pairs,  $n_1$  and  $n_2$  are the bunch populations (protons per bunch) in beam 1 and beam 2 respectively (together forming the bunch charged product),  $f_r$  is the machine revolution frequency, and  $\Sigma_x$  and  $\Sigma_y$  are the width and the height of the proton beams.

The number of protons per bunch, the number of bunches per beam, and the revolution frequency are all set by the beam operators. The widths of the proton beams are measured in a process known as a Van der Meer (*vdm*) scan [16]. In a *vdm* scan, the beams are separated by steps of a known distance. The collision rate is measured as a function of this separation, and the width of a gaussian fit to the distributions yields the width of the beams in the direction of the separation.

The total integrated luminosities provided by the LHC and recorded by

ATLAS in 2011 are shown in Figure 3.3. These events form the dataset analyzed in this thesis. By means of the beam-separation or  $vM$  scans, as well as other techniques to measure the bunch charged product, the ATLAS Collaboration has determined that the uncertainty on its luminosity measurement is  $\delta L = \pm 3.7\%$ . For a complete description of the methods used and the systematic errors evaluated see reference [15].

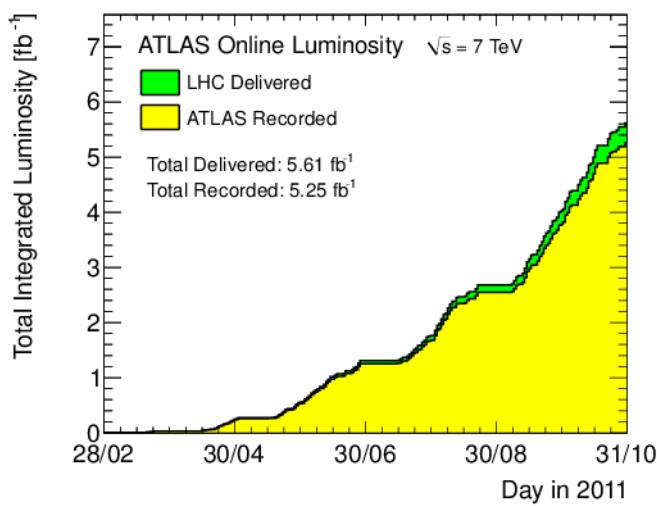


Figure 3.3: Total luminosity delivered by the LHC and recorded by ATLAS during the 2011  $\sqrt{s} = 7$  TeV proton-proton run

As anticipated, due to the cross-section for interaction and the number of protons per bunch, the possibility to observe multiple  $pp$  interactions per bunch crossing increases proportionally. This phenomenon, referred to as “pile-up”, can really occur in two distinct forms. The first form is the presence of multiple  $pp$  collisions (different from the interaction of interest) in the same bunch crossing, referred to as “in-time” pile-up. The second form of pile-up takes place due to electronic integration times within the detector. Certain detector components are actually sensitive to multiple bunch cross-

ings due to the long electronic signals generated in the response to energy depositions or charge collection. One or more  $pp$  collisions in a bunch-crossing different from that which produced the collision of interest can then affect the measurement. This form of pile-up is referred to as “out-of-time” pile-up and will become more and more important as the LHC bunch spacing gets closer to the nominal value, 25 ns.

The fraction of events with pile-up increased significantly since the data taking started. The experimental signature of this fact is obtain via the number of reconstructed primary vertices, or NPV. The effect of the event NPV is an important concern for the measurement of jet properties and will be discussed in the next chapters.

## 3.2 The ATLAS Detector

The ATLAS detector [17] is one of the two general purpose particle detectors built for probing  $pp$  collisions at the LHC. As it was described in the previous section, inside the LHC, bunches of up to  $10^{11}$  protons will collide 40 million times per second to provide 14 TeV proton-proton collisions at a nominal luminosity of  $10^{34}\text{cm}^{-2}\text{s}^{-1}$ . These high interaction rates and energies, as well as the requirements for high precision physics measurements set the standars for the design of the detector. At even 7 TeV center-of-mass energy, the LHC interactions result in high particle multiplicity, requiring fine detector granularity, and particle production at forward rapidity, requiring large detector angular coverage.

To achieve these performance goals, a design consisting of multiple detector sub-systems with cylindrical symmetry around the incoming beams is used as shown in Fig. 3.4. Closest to the interaction point the inner tracking

detector is placed, providing charged particle reconstruction. The magnet configuration comprises a thin superconducting solenoid surrounding the inner detector cavity, and three large superconducting toroids (one barrel and two end-caps) arranged with an eight-fold azimuthal symmetry around the calorimeters. This fundamental choice has driven the design and size (44 m in length and 25 m in height) of the rest of the detector. Outside the solenoid, a calorimeter system performs electron, photon, tau, and jet energy measurements. Finally, the calorimeter is surrounded by the muon spectrometer where an array of muon drift chambers perform muon identification and momentum measurements.

The ATLAS detector coordinate system is used to describe the position of particles as they traverse these subdetectors. It is a right-handed coordinate system, with  $z$  pointing along the beam direction, positive  $x$  pointing toward the center of the LHC ring, and positive  $y$  pointing up. The  $x-y$  plane is referred to as the transverse plane, and the  $z$  direction as the longitudinal direction. The azimuthal angle  $\phi$  is measured as usual around the beam axis, and the polar angle  $\theta$  is the angle from the beam axis. The pseudorapidity is defined as  $\eta = \ln \tan(\frac{\theta}{2})$ , regions of low  $\eta$  are referred to as “central”, and regions of high  $\eta$  are referred to as “forward”. The transverse momentum  $p_T$  is defined in the  $x-y$  plane unless stated otherwise. The distance  $\Delta R$  in the pseudorapidity-azimuthal angle space is defined as  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ .

To meet the extremely high demands that the LHC luminosity places on the speed with which ATLAS must record data, a dedicated trigger and data acquisition (TDAQ) system is used. The interaction rate at the design luminosity is approximately 1 GHz, while the event data recording, based on technology and resource limitations, is limited to about 200 Hz. This requires a high rejection of minimum-bias???? processes while maintaining maximum

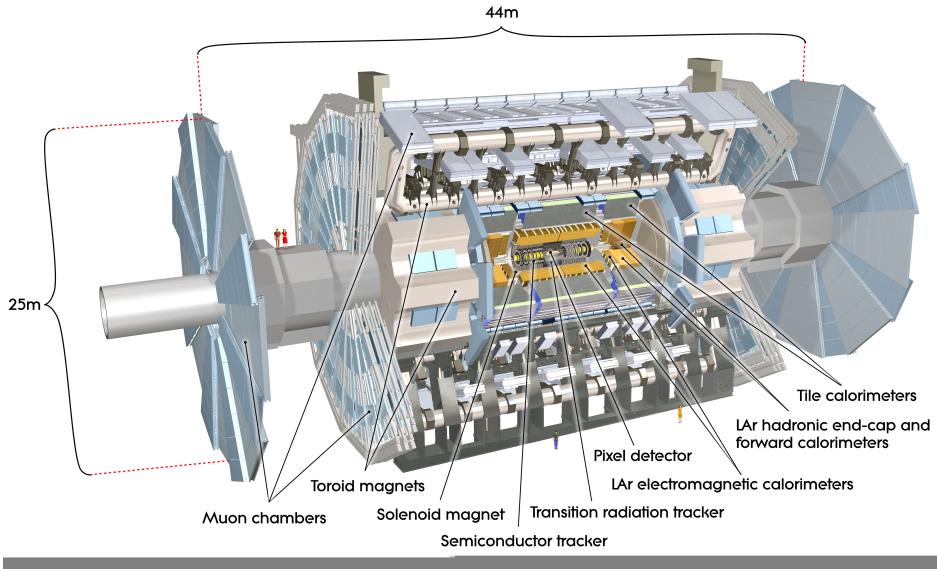


Figure 3.4: El detector de ATLAS

efficiency for the new physics. The Level-1 (L1) trigger system uses a subset of the total detector information to make a decision on whether or not to continue processing an event, reducing the data rate to approximately 75 kHz (limited by the bandwidth of the readout system, which is upgradeable to 100 kHz). The subsequent two levels, collectively known as the high-level trigger (HLT), are the Level-2 (L2) trigger and the event filter. They provide the reduction to a final data-taking rate of approximately 200 Hz.

### 3.2.1 Inner tracking system

The inner tracking system or inner detector (ID) is composed of three sub-detectors: the pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). The goal of these three is to provide charged particle trajectory reconstruction and momentum measurements with an

overall acceptance in pseudorapidity of  $|\eta| < 2.5$  and full  $\phi$  coverage.

The sensors which built this system register signals, referred to as “hits”, in response to the passage of charged particles. The ID is immersed in a 2 T magnetic field, generated by the central solenoid. The positions of the registered hits are combined to form tracks, with the radius of curvature of the tracks (caused by the presence of the magnetic field) providing a measurement of the particles transverse momentum. The track reconstruction efficiency ranges from 78% at  $p_T^{track} = 500$  MeV to more than 85% above 10 GeV, averaged across the full  $\eta$  coverage [18]. The transverse momentum resolution of  $\sigma_{p_T}/p_T = 0.05$  [19] (upper bound) and a transverse impact parameter resolution of  $\sim 20 \mu m$  for high momentum resolution particles in the central  $\eta$  region[20].

The pixel detector, SCT, and TRT sensors are arranged on concentric cylinders around the beam axis, known as barrel layers, and on disks perpendicular to the beam at either end of the barrel, known as end-caps. A more complete description of these systems is given below. The overall layout of the inner detector is shown in Fig. 3.5.

### The Pixel detector

The pixel detector consists of three concentric barrel layers. The innermost one, the so called “b-layer” due to its role in identifying  $b$ -quarks initiated jets, is located at 5 cm from the interaction region. Three additional disks are located at each end-cap, producing typically three pixel position measurements per charged particle track. Each layer or disk is instrumented with modules that form the basic unit of data acquisition, each with 47232 pixels. All pixel sensors are identical and have a minimum pixel size in  $r - \phi \times z$  of  $50 \times 400 \mu m^2$ . The intrinsic accuracies in the barrel are  $10 \mu m$

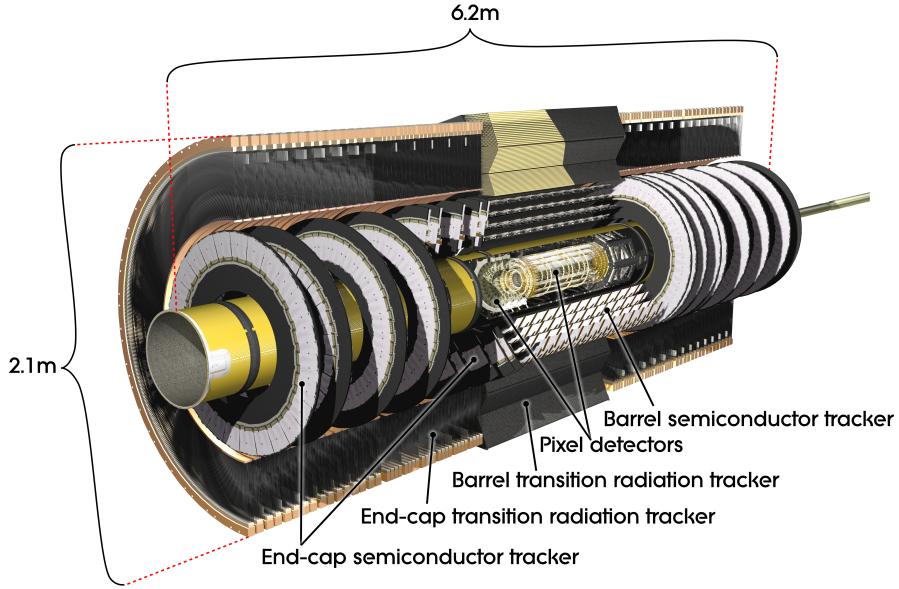


Figure 3.5: Layout of the ATLAS inner detector

in  $r - \phi$  and  $115 \mu\text{m}$  along  $z$ , or along  $r$  in the end-caps. The pixel detector has approximately 80.4 million readout channels, an order of magnitude more readout channels than the rest of ATLAS combined, and it extends to a total length of  $z \sim \pm 650 \text{ mm}$  and radius of  $r \sim 150 \text{ mm}$ , providing good reconstruction efficiency for tracks up to  $|\eta| < 2.5$ .

### The SCT

The SCT consists of four barrel layers and nine end-cap layers surrounding the pixel detector, resulting in at least four hits along every charged particle track. The SCT barrel reaches to  $z \sim \pm 750 \text{ mm}$  and  $r \sim 515 \text{ mm}$ , while the end-cap covers out to  $z \sim \pm 2720 \text{ mm}$  and  $r = 560 \text{ mm}$ . There are 15,912 SCT module sensors, each 12.8 cm long and approximately  $285 \mu\text{m}$  thick.

In the barrel region, these modules use small-angle (40 mrad) stereo strips

to measure both coordinates, with one set of strips in each layer parallel to the beam direction, measuring the  $\phi$  coordinate directly . In the end-cap region, the detectors have a set of strips running radially and a set of stereo strips at an angle of 40 mrad. The mean pitch of the strips is  $80 \mu\text{m}$ . The intrinsic accuracies per module in the barrel are  $17 \mu\text{m}$  in  $r - \phi$  and  $580 \mu\text{m}$  in  $z$  (or  $r$  in the end-caps). The total number of readout channels in the SCT is approximately 6.3 million. A hit is registered only if the pulse height in a channel exceeds a preset threshold ( $\sim 1 \text{ fC}$ ). The charged measured in the strip is then recorded into a memory buffer that is only read out and used for tracking if a trigger is received signaling that the event should be considered in more detail.

## The TRT

The TRT surrounds the silicon detectors and is comprised of up to 76 layers of longitudinal straw tubes in the barrel, extending to  $z \sim \pm 710 \text{ mm}$  and  $r \sim 1060 \text{ mm}$ , and 160 radial straw planes in each end-cap cylinders, reaching  $z \sim \pm 2710 \text{ mm}$  and  $r \sim 1000 \text{ mm}$ .

The TRT sensors are thin drift tubes consisting of cathode metal straws filled with an ionizing gas mixture of xenon, oxygen, and  $\text{CO}_2$ , with an anode wire running down the center of the straw. The passage of a charged particle through the gas produces positive ions and free electrons, which travel to the cathode and anode, respectively, under the influence of an applied voltage of 1600 V. Comparing the time that the signals are received at the cathode and the anode gives a drift time measurement that can be used to calculate the impact parameter of the particle. This method gives no information on the position along the length of the straw.

To give the best resolution of particle trajectories as they bend in the

solenoidal field, the straws lie along the beam direction in the barrel and radially in the end-caps. The straw diameter of 4 mm causes a maximum drift time of approximately 48 ns and an intrinsic accuracy of 130 m along the radius of the straw.

In addition to directly detecting charged particles produced by the collision, the TRT also measures the transition radiation induced by the passage of these particles through polypropylene sheets placed between the drift tube straws. Transition radiation refers to the photons emitted by charged particles as they pass from one material into another with a different dielectric constant. These photons yield a much larger signal amplitude than the charged particles, so separate thresholds in the electronics can be used to distinguish the two.

One of the most important tasks of the inner detector for is to provide accurate collision vertex identification, exploiting the excellent position resolution and tracking efficiency. Vertices are reconstructed by matching inner detector tracks with  $p_T > 150$  MeV back to a common origin.

### 3.2.2 The Calorimeter System

The purpose of the ATLAS calorimeter system is to measure the energy of electrons, photons, taus and jets, within the pseudorapidity region of  $|\eta| < 4.9$  and with full  $\phi$  symmetry and coverage around the beam axis. It also provides fast position and energy measurements to serve as trigger signals for these objects as well as the missing transverse energy.

It consists of an electromagnetic (EM) calorimeter system and a hadronic one. Both calorimeters are sampling calorimeters meaning that they utilize alternating layers of absorber material, composed of heavy atoms that interact with energetic particles and cause them to loose energy, and an active

material, that produce a signal in response to the deposited energy.

It is very important to note that each calorimeter is segmented both laterally and longitudinally. This leads to high granularity EM calorimeter that maximizes the accuracy with which electrons and photons can be reconstructed. It also provides detailed information on the transverse and longitudinal shower shapes of hadronic jets.

The calorimeters closest to the beam-line are housed in three cryostats, one barrel and two end-caps. The barrel cryostat contains the electromagnetic barrel calorimeter, whereas the two end-cap cryostats each contain an electromagnetic end-cap calorimeter (EMEC), a hadronic end-cap calorimeter (HEC), located behind the EMEC, and a forward calorimeter (FCal) to cover the region closest to the beam. These calorimeters use liquid argon as the active detector medium and need to be maintained at a constant temperature of  $\sim 88\text{K}$ . Liquid argon (LAr) has been chosen for its intrinsic linear behaviour (production of ionization charge as a function of incident charge), its stability of response over time and its intrinsic radiation-hardness. An illustration of these components together with the tile hadronic calorimeter can be found in Fig. 3.6. Further specifications are given in the next sections.

### **The electromagnetic calorimeter**

The EM calorimeter uses lead as the absorber and LAr as the active material.

Con el fin de asegurar una perfecta cobertura en todo el rango de pseudo-  
rapidez y proveer una completa simetría en la coordenada  $\phi$  se ha elegido una  
geometría en forma de acordeón. En la región dedicada a física de precisión  
( $|\eta| < 2.5$ ) el calorímetro electromagnético está dividido en cuatro secciones  
de muestreo. La primera o *Presampler* consiste en una delgada capa de argón  
desprovista de material absorbente, cuyo propósito es la corrección por la

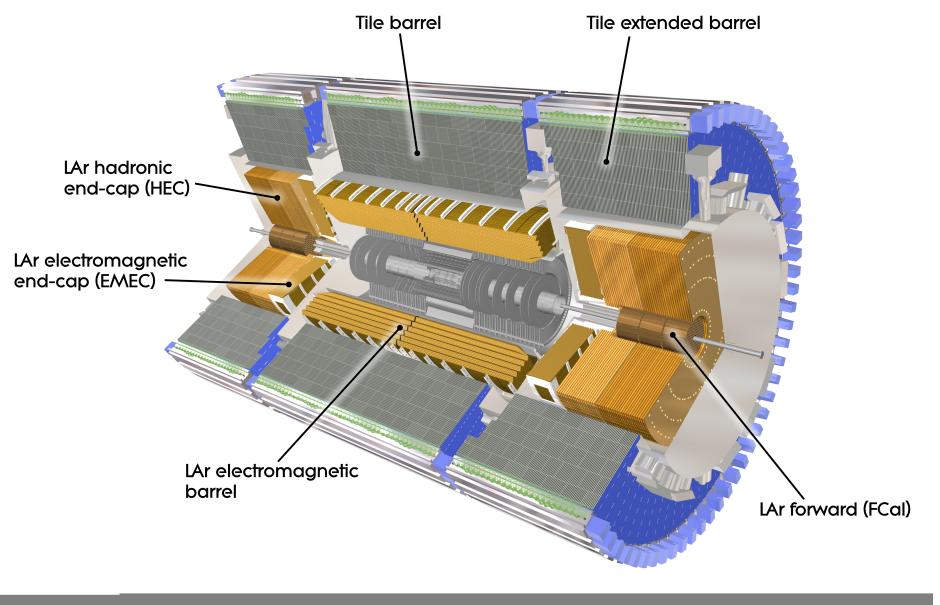


Figure 3.6: Layout of the ATLAS electromagnetic and hadronic calorimeter systems. The total length is  $\sim 12$  m, extending to a maximum radius of 4.25 m.

pérdida de energía en el solenoide y las paredes del crióstato. La segunda sección posee una profundidad de  $4.3X_0$ <sup>1</sup> y en ella la lectura se lleva a cabo mediante celdas en forma de tiras delgadas en  $\eta$ , dando una buena resolución en dicha coordenada, con  $\Delta\eta = 0.0031$ . La sección de *2nd Sampling* ( $16X_0$ ) es donde se deposita la mayoría de la energía, teniendo ambas coordenadas igual importancia. Allí el tamaño de celda es de  $\Delta\phi\Delta\eta = 0.0245 \times 0.0245$ . Sólo los electrones más energéticos llegarán a la cuarta sección de muestreo (*3rd Sampling*).

Las ruedas calorimétricas comienzan en  $|\eta| = 1.5$  y continúan abajo hasta  $|\eta| = 3.2$ , pero con un tamaño de celda mayor por encima de  $|\eta| = 2.5$ .

En la región de bajo ángulo, el calorímetro hadrónico funciona también con argón líquido de manera de resistir los altos niveles de radiación. Su diseño es más sencillo que el del calorímetro electromagnético y como absorbentes posee placas paralelas de cobre, perpendiculares al haz. La región de esta parte del calorímetro hadrónico, que cubre hasta  $|\eta| = 4.9$ , está hecha de cobre/tungsteno. La elección de este material es necesaria para limitar el ancho y la profundidad de las lluvias provenientes de jets de altas energías cercanas a la línea del haz.

### The hadronic calorimeter

El calorímetro hadrónico de tejas de ATLAS está compuesto por tres barriles, uno central de 5,6m y dos extensiones de 2,9m cada una. El radio interno es de 2,2m y el externo, de 4,2m.

---

<sup>1</sup> $X_0$  es el símbolo utilizado para la Longitud de Radiación (*Radiation length*), definida como la distancia media sobre la cual un electrón muy energético pierde  $1/e$  de su energía por bremsstrahlung o bien, como  $7/9$  del camino libre medio en la producción de pares. Constituye una escala de longitud apropiada para describir cascadas electromagnéticas de alta energía.

Cada barril está dividido en 64 cuñas azimutales o módulos, con una estructura periódica en la dirección paralela al haz. Cada módulo es una estructura de tejas de hierro (material absorbente) alternadas con tejas de plástico centellador, dispuestas en un plano paralelo al eje del haz. Los materiales centelladores emiten luz en forma de pequeños pulsos cuando son atravesados por partículas o radiación; acoplando el centellador a un fotomultiplicador el pulso de luz se convierte en un pulso eléctrico que puede ser analizado. En el caso del calorímetro hadrónico de ATLAS, cuando una partícula atraviesa una teja centelladora emite luz en el rango del ultravioleta, de intensidad proporcional a la energía depositada por la partícula.

Se ha elegido una segmentación proyectiva en torres de  $\Delta\phi\Delta\eta = 0.1 \times 0.1$  y cada una de estas torres está dividida, en profundidad, en tres celdas, leídas individualmente por dos fotomultiplicadores para conseguir redundancia en la señal. La luz generada en las tejas es recogida mediante fibras ópticas que cambian la longitud de onda y transportada a los fotomultiplicadores (este subdetector posee unos 10.000 fotomultiplicadores).

El calorímetro hadrónico debe tener el espesor suficiente para contener la energía de los hadrones. Con este propósito y para obtener una buena resolución se han elegido 11 longitudes de absorción como camino previo a las cámaras de muones.

The ATLAS hadronic calorimeters are non-compensating calorimeters, meaning that the signal response for incident electromagnetic and hadronic particles is different. This difference for hadronic objects, such as jets, must be accounted for when reconstructing the final object energy in a separate step after the signal reconstruction. The procedure for translating the measured EM energy to the correct hadronic scale is detailed in Section ??.

### 3.2.3 The Muon System

El sistema de muones sirve a un doble propósito: funciona como sistema de disparo (o *trigger*) para la selección de eventos con muones de alta energía, y como espectrómetro de muones de alta precisión. En este sentido, este detector llevará a cabo la identificación de los muones producidos en las colisiones  $p$ - $p$ , determinando sus trayectorias y momentos. El sistema consiste en un conjunto de toroides (llamamos así, por su forma, a los tres conjuntos de bobinas que proveen el campo magnético toroidal) y cámaras de tubos de deriva que se encuentran rodeando al calorímetro. En la parte del barril del detector, las cámaras están situadas en el interior del toroide lo que permite la medición del momento de las partículas a partir de la desviación de sus trayectorias en el campo magnético. En las tapas, donde la presencia del crióstato impide posicionar las cámaras dentro del campo magnético, el momento es medido a partir de la diferencia entre los ángulos de entrada y salida del imán. En el plano trasversal, tanto en la región del barril como en las tapas laterales, el sistema de cámaras estará dividido en 16 sectores, siguiendo la simetría determinada por las 8 bobinas del barril central del sistema magnético. Las cámaras cubren el espacio entre las bobinas, y todo el rango acimutal en la región que las rodea. Los sectores se numeran comenzando a partir de  $\phi = 0$ , en el sentido contrario de las agujas del reloj, teniendo en la dirección vertical a los sectores 6 (en la parte superior del detector) y 13 (sector inferior).

Los cámaras de tubos de deriva (MDTs) son cámaras proporcionales hechas de tubos de aluminio de 30 mm de diámetro y longitudes variables de 70 a 630 cm, con un hilo central de  $50\mu\text{m}$  de diámetro, de W-Re. En la región del barril dichas cámaras están distribuidas en 3 capas cilíndricas concéntricas (estaciones) alrededor del haz, de 5; 7,5 y 10 metros de radio.

Los tubos están dispuestos de manera transversal al eje  $z$  de manera de medir la coordenada en el plano de desviación de la trayectoria de la partícula (plano  $Rz$ ). Estas cámaras miden el tiempo de deriva de la ionización producida por el paso del muón, teniendo una resolución de  $80 \mu\text{m}$ .

Cada cámara MDT está cubierta por una o dos cámaras de placas resistentes (RPCs). Cada una de ellas encierra un volumen de gas entre planchas resistentes de baquelita, dotada una de ellas con tiras de electrodos. Dado que los tubos de deriva poseen un diámetro relativamente grande que resulta en un tiempo de deriva máximo de 480ns, mucho mayor que los 25 ns entre cruce de *bunches*, se requieren cámaras especiales de disparo para la selección de eventos. La función de trigger en el barril es provista por tres capas de RPCs, situadas, dos de ellas, a ambos lados de la segunda estación de MDTs y la restante, en la cara interior de la estación más externa. En las tapas, esta función es cumplida por tres estaciones de TGCs (*Thing Gap Chambers*). Estas cámaras son similares en diseño a cámaras proporcionales multihilo, con la diferencia de que poseen una distancia cátodo-cátodo menor que la pendiente del ánodo (hilo). Las cámaras de disparo proveen una estimación de las coordenadas  $\phi$  y  $\eta$  del punto de impacto de la traza, mientras que las cámaras MDTs darán (con mayor precisión) la coordenada  $\eta$ .

En la región de bajo ángulo, donde la densidad de trazas es mayor, se utilizan cámaras de tiras de cátodos (CSCs) de granularidad más fina comparadas con las MDTs, para la detección de trayectorias. Estas cámaras son cámaras proporcionales, con un espacio entre hilo de 2,5 mm. Cada una de ellas proporciona medida de dos coordenadas y puede operar en condiciones de alto campo magnético.

En la figura 3.7 se puede ver un esquema del especlómetro de muones, donde se indica la posición de las diferentes cámaras descriptas.

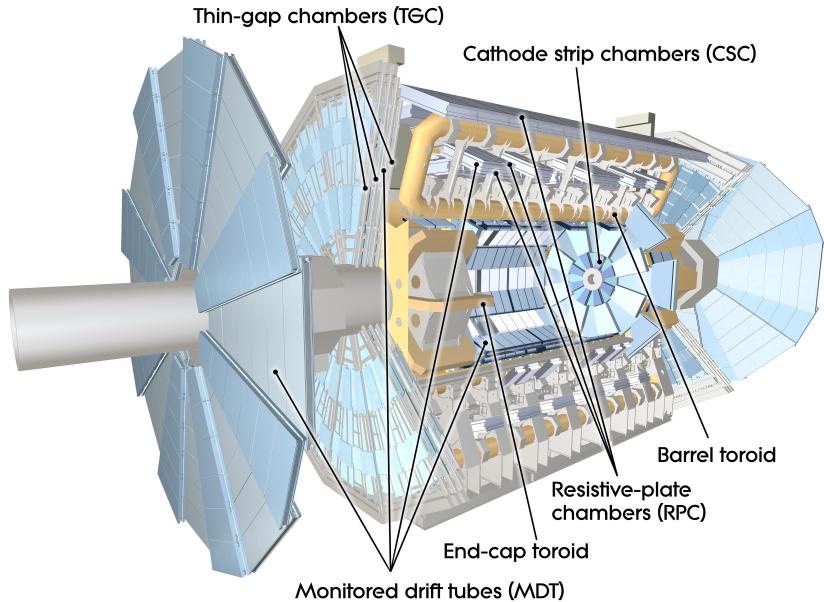


Figure 3.7: Muon Chamber

El sistema de imanes superconductores de ATLAS consiste en un solenoide central que provee el campo magnético necesario al detector interno, rodeado por un arreglo de bobinas o bucles, con forma de pista de carrera, que generan un campo magnético toroidal para el espectrómetro de muones. Todo el sistema es enfriado de manera indirecta mediante el flujo de helio líquido a 4,5 K.

El solenoide es un electroimán superconductor de 5,3 m de largo, situado en el interior del calorímetro electromagnético. Comparte el crióstato con el calorímetro de argón líquido, evitando la presencia de dos paredes criostáticas y reduciendo así la cantidad de material introducido. La longitud del solenoide es considerablemente más pequeña que la del barril del detector de trazas. Este es el resultado de un compromiso: un bobinado corto reduce la cantidad de material introducido mientras que uno largo pro-

porciona un campo magnético más uniforme en dicho detector. El campo magnético a lo largo del eje  $z$  es de 2 T en el punto de interacción.

El arreglo de bobinas está dividido en un barril central y dos regiones laterales, al igual que los detectores. El barril central está constituido por 8 bobinas de 5 m de ancho por 25 metros de largo aproximadamente, dispuestas simétricamente alrededor del haz de manera radial. Las bobinas del barril se encuentran en crióstatos separados, mientras que las 8 bobinas en cada una de las tapas o toroides laterales están ubicadas en un crióstato común.

Con un campo magnético toroidal las partículas atravesarán todo el rango de pseudorapidez casi perpendicularmente al haz. El número pequeño de bobinas que generan el campo toroidal resulta en una intensidad de campo que varía fuertemente con la coordenada  $\phi$ . En el barril el campo magnético es de 2 T, mientras que en las tapas es de 4 T en las zonas de mayor intensidad.

### 3.2.4 Trigger and Data Adquisition

En este capítulo se analiza la estructura del trigger de ATLAS, y el sistema de adquisición y flujo de los datos. Se presenta, asimismo, una breve descripción de los algoritmos usados en la reconstrucción de trayectorias para la selección de eventos en el detector interno.

#### Arquitectura general

El sistema de Trigger y Adquisición de Datos[?] de ATLAS está basado en tres niveles de selección *online*: Nivel 1, Nivel 2 y Filtro de Eventos. Cada nivel es más lento pero más preciso que el anterior. Trabajando con una frecuencia de interacción de  $10^9$  Hz y luminosidades del orden de  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , este sistema será el encargado de reducir la frecuencia de eventos inicial

de 40 MHz a 200Hz, que es la velocidad con la que pueden almacenarse.

En la figura 3.8 se muestra un vista simplificada de los principales componentes y funciones.

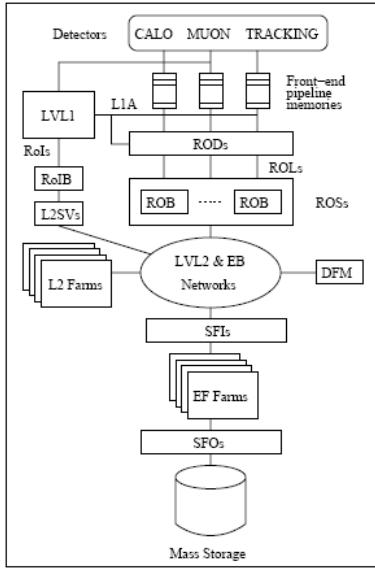


Figure 3.8: Principales componentes del sistema de trigger y adquisición de datos de ATLAS.

El mecanismo que lleva a cabo el movimiento de la información (Data Flow System), es el responsable de recibir los datos de los detectores, pasando parte de ellos al sistema de trigger y enviando luego, los eventos seleccionados al lugar de almacenamiento. Siguiendo el esquema de la figura, la comunicación entre los *drivers* de lectura de cada detector (RODs) y el sistema de adquisición de datos, está dada por los *buffers* de almacenamiento transitorio (ROBs). La información de los eventos aceptados por el Nivel 1 son transportados de los primeros al sistema de lectura (ROS), que consta de numerosos ROBs, guardando los datos a la espera de la decisión del trigger. La información requerida por el segundo nivel es provista por estos últimos.

Los eventos aceptados son reconstruidos (a partir de fragmentos contenidos en diferentes ROBs) y pasados al siguiente nivel.

El Nivel 2 y el Filtro de Eventos componen el *High-level Trigger* (HLT) de ATLAS. El Nivel 2 trabaja a la frecuencia de aceptación del Nivel 1, utilizando una secuencia de rápidos algoritmos de selección que operan típicamente sobre una fracción de los datos del evento, contenida en regiones del detector previamente seleccionadas por ese nivel (ver el mecanismo de la región de interés en la siguiente sección). Si la decisión del Nivel 2 es rechazar el evento, los datos del mismo son eliminados de los buffers correspondientes. Si el evento es aceptado, se reconstruye en el EB (Event Builder) y es pasado al Filtro de Eventos. Este nivel ejecutará algoritmos de reconstrucción más sofisticados, adaptados de aquellos para el análisis *offline*, utilizando información detallada de los detectores para efectuar el proceso de selección final, que determinará cuáles son los eventos que serán guardados para posteriores estudios.

En las siguientes secciones se presenta una descripción más detallada de los niveles de trigger.

## El Nivel 1

El primer nivel de trigger de ATLAS es implementado mediante hardware. Éste realiza una decisión inicial a partir de la información provista por los calorímetros y del detector de muones, basando su estrategia en la combinación de objetos en coincidencia.

En el sistema de muones, los candidatos de alto momento transverso son identificados en las cámaras especiales de trigger: RPCs en el barril y TGCs en las tapas. En el caso del calorímetro, se definen una serie de conjuntos de umbrales de  $p_T$  para cada objeto (electrones, fotones, jets, etc.),

seleccionando aquellos que pasen los criterios de selección correspondientes al evento físico de interés.

Puesto que la decisión de aceptar un evento no puede ser realizada en los 25 ns que median entre dos cruces de *bunches*, los subdetectores almacenan localmente la información del mismo en *pipelined buffers* hasta que el Nivel 1 efectúa la selección. Luego, los datos son enviados a los RODs específicos de cada detector para luego dirigirse a los ROBs, donde son almacenados hasta que la decisión del Nivel 2 sea alcanzada. Cuando un evento es aceptado, el Nivel 1 comunica la decisión al mecanismo que se encargará de construir una Región de Interés (RoI). Este mecanismo es una importante pieza sobre la que descansa la estrategia del sistema de trigger; a través del mismo, el Nivel 2 hará uso de la información del evento en regiones localizadas del detector, de manera que los algoritmos de reconstrucción en ese nivel sólo transfieran los ROBs necesarios para arribar a una rápida decisión. La RoI contendrá la información de la posición ( $\eta$  y  $\phi$ ) y el momento de los objetos candidatos.

Este nivel está diseñado para llevar a cabo su decisión en un tiempo menor a  $2.5 \mu\text{s}$ , medidos desde la colisión  $p\text{-}p$ , hasta que la información del evento está disponible en la electrónica de salida de los detectores. En este proceso la frecuencia de eventos será reducida a 75KHz (límite fijado por la electrónica).

### 3.2.5 El HLT

El High-level Trigger de ATLAS abarca la segunda y tercera etapa de la selección de eventos. Comprende el Nivel 2 y el Filtro de Eventos, y contiene además, el Software de Selección (ESS). Este último comparte la estructura usada por el Offline para los códigos de selección, facilitando el análisis *offline* de los datos, y el desarrollo de algoritmos en el HLT.

El punto de entrada del trigger es el resultado del Nivel 1. Éste provee información acerca de la región de interés, fundamental para el rápido funcionamiento de los algoritmos del Nivel 2. Así, los datos del Nivel 1 guían la selección del Nivel 2; y ésta a su vez guiará la del Filtro de eventos, como se ilustra en la figura 3.9.

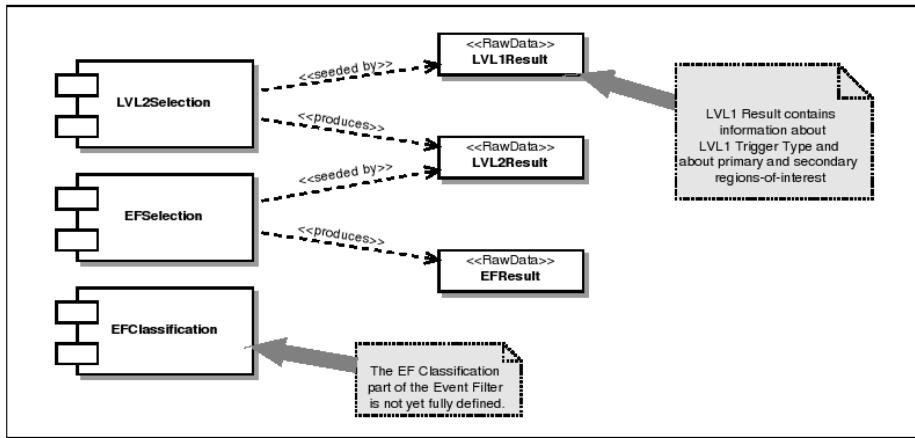


Figure 3.9: Cadena de selección del *High-level Trigger* de ATLAS. Cada nivel es guiado por el resultado del paso anterior.

## El Nivel 2

La tarea específica del Nivel 2 es reducir la frecuencia de eventos de  $\sim 100$  kHz a alrededor de 2 kHz, combinando la información de todos los detectores para su decisión global. A diferencia del Nivel 1, esta segunda etapa de selección realiza operaciones no sincronizadas sobre los eventos, con un tiempo de decisión de 10 ms.

El Nivel 2 utiliza las regiones de interés provistas por el Nivel 1. Cada región es examinada en el subdetector de origen (calorímetro o sistema de muones) para su confirmación; para luego buscar información de otros sub-

detectores. En el caso del trigger de muones, el poder de rechazo del Nivel 2 proviene de ajustar los umbrales de  $p_T$ , respecto de los utilizados en el primer nivel, a partir de la información de las cámaras de precisión del sistema de muones (MDTs) y la correspondiente al detector interno. Los procesadores del Nivel 2 son los encargados de ejecutar luego el software de selección de eventos, utilizando la información almacenada en los *buffers*. Usando las RoIs del Nivel 1, el Nivel 2 accederá de manera selectiva a los datos en los ROBs, moviendo sólo la información requerida para efectuar la decisión. Típicamente, sólo una pequeña fracción del detector, correspondiente a las regiones centradas en los objetos indicados por el Nivel 1, serán necesitados por el segundo nivel.

Hasta que un evento es aceptado o rechazado (en  $\sim 10$  ms), los datos son retenidos en los ROBs. En caso de aceptación, los fragmentos del evento almacenados en distintos buffers serán requeridos por el sistema de control del Nivel 2 (L2SVs) para ser enviados al constructor de eventos (EB). El evento ensamblado es guardado en una única dirección de memoria para ser utilizado por el Filtro de Eventos. El tamaño promedio de un evento será del orden de 1,5 MB.

### **El Filtro de Eventos**

Luego del Nivel 2, la última etapa de selección *online* es realizada por el Filtro de Eventos (EF). El EF emplea algoritmos y métodos similares a los implementados en el análisis *offline*, adaptados para su corrida en el tiempo real del experimento; su poder de rechazo radica en el uso de algoritmos y criterios de selección más complejos, que por límites en el tiempo de procesamiento no pueden ser utilizados en el Nivel 2.

El EF utilizará información actualizada de la calibración y alineamiento

del detector y un completo mapa del campo magnético; llevando a cabo con ello la selección final del evento físico que será guardado para su estudio en el Offline. La frecuencia de aceptación del nivel anterior será reducida en un orden de magnitud, almacenando a una tasa de  $\sim 100$  MB/s.

### El software de selección

La tarea del software de selección (ESS) es la selección y clasificación de los eventos. Candidatos tales como electrones, jets, muones, etc., representados por objetos abstractos, son reconstruidos utilizando un particular conjunto de algoritmos. Un evento es seleccionado si el objeto reconstruido satisface al menos una de las signaturas establecidas en el menú del sistema de disparo. En el Nivel 2 y el Filtro de Eventos (EF), los eventos serán rechazados si no pasan los específicos criterios de selección, diseñados para la reducción de la frecuencia de eventos, al límite dado por la velocidad a la que éstos pueden ser almacenados.

El ESS se compone de una infraestructura y un conjunto de programas de selección para las dos etapas del HLT. Los algoritmos de reconstrucción para el trigger están basados en aquellos utilizados para la reconstrucción *offline*, pero correrán *online* en el entorno de software provisto por los procesadores del Nivel 2 y el EF.

De manera de facilitar el desarrollo de los algoritmos del HLT y simplificar los estudios del Offline; el ESS ha sido diseñado de manera de poder ser ejecutado directamente en el entorno provisto por la estructura de software de análisis offline del experimento, ATHENA[?]. La estructura dada por este paquete de software es lo suficientemente flexible como para abarcar una variedad de procesos, incluyendo no sólo algoritmos de trigger sino también tareas de calibración y monitoreo. Se ha destinado un apéndice (A) para su

descripción.

En el Offline, la tarea del ESS es la de emular la cadena completa de selección *online*. Para su ejecución el sistema se sirve de cuatro sub-paquetes: el direccionamiento o *Steering*, los algoritmos del HLT, y los paquetes de software para la clasificación y movimiento de los datos, EDM (*Event Data Model*) y el DM (*Data Manager*). Los últimos toman los datos del evento en el formato que poseen a la salida de los sistemas de lectura (*Raw data* en formato *byte stream*), y los convierten en objetos que puedan ser usados por los algoritmos en la cadena de selección (*Raw Data Objects*).

La tarea de los algoritmos del HLT es la de analizar los datos del evento, reconstruyendo partes del mismo, luego de la selección del Nivel 1. El paquete se compone de dos subconjuntos principales:

- Programas de preparación de datos. Son los algoritmos ejecutados por los sistemas EDM y DM para la conversión del formato de los datos del evento.
- Algoritmos FEX o de *Feature Extraction*. Comprende los programas de reconstrucción y los llamados algoritmos de “hipótesis”. Estos últimos (a los primeros nos referiremos en la siguiente sección) son aquellos programas que se encargan de eliminar, una vez realizada la reconstrucción, aquellos candidatos que no cumplen con las características o atributos asignados al evento físico en consideración (hipótesis), aplicando específicos criterios de selección. La presencia de los algoritmos de hipótesis es fundamental en la secuencia del HLT ya que evita la ejecución innecesaria de algoritmos al descartar eventos en las primeras etapas de la cadena.

Por último, el subpaquete de *Steering* es aquel que organiza el proce-

samiento de los datos del evento en el Nivel 2 y el Filtro de eventos; controlando el orden en el que los algoritmos de reconstrucción e hipótesis son ejecutados. El Steering define la secuencia del HLT, y manipula los resultados en cada paso de selección de manera que la decisión del trigger sea alcanzada.

### **3.2.6 Data quality**

### **3.2.7 Simulation of particle interactions in the ATLAS Detector**

# Chapter 4

## Jet reconstruction and *b*-Tagging

Experimental data and simulated events were both reconstructed using the latest version available of the ATLAS software. In this chapter we briefly describe the reconstruction of the two key objects used in this analysis, namely jets and tracks. We describe how *b*-Tagging is done in ATLAS.

### 4.1 Jets reconstruction and calibration

Jets are reconstructed using the anti- $k_t$  jet algorithm [21] with a distance parameter  $R = 0.4$ , using calorimeter topological clusters [22] as input. Several quality criteria are applied to eliminate “fake” jets that are caused by noise bursts in the calorimeters and energy depositions belonging to a previous bunch crossing [23].

The jet energies are corrected for inhomogeneities and for the non-compensating nature of the calorimeter by using  $p_T$  – and  $\eta$ -dependent calibration factors determined from Monte Carlo simulation [24]. This calibration is referred to

as the EM+JES scale. Using test beam results, in-situ track and calorimeter measurements, estimations of pile-up energy depositions, and detailed Monte Carlo comparisons, an uncertainty on the absolute jet energy scale was established. This uncertainty is smaller than  $\pm 10\%$  for  $\eta < 2.8$  and  $p_T > 20$  GeV. More sophisticated techniques undergoing commissioning, such as local cluster weighting, are expected to considerably improve the jet energy uncertainty and resolution [25].

## 4.2 *b*-jet Tagging

Jets are classified as *b*-quark candidates by the ATLAS MV1 *b*-tagging algorithm, based on a neural network that combines the information from three high-performance taggers: IP3D, SV1 and JetFitter [3]. These three tagging algorithms use a likelihood ratio technique in which input variables are compared to smoothed normalized distributions for both the *b*- and background (light- or in some cases *c*-jet) hypotheses, obtained from Monte Carlo simulation. The IP3D tagger takes advantage of the signed transverse and longitudinal impact parameter significances. The SV1 tagger reconstructs an inclusive vertex formed by the decay products of the *b*-hadron and relies on the invariant mass of all tracks associated to the vertex, the ratio of the sum of the energies of the tracks in the vertex to the sum of the energies of all tracks in the jet and the number of two-track vertices. The JetFitter tagger exploits the topology of the primary, *b*- and *c*-vertices and combines vertex variables with the flight length significance. The *b*-tagging performance is determined using a simulated  $t\bar{t}$  sample and is calibrated using experimental data with jets containing muons and with a sample of  $t\bar{t}$  events [26].

# Chapter 5

## Double $B$ -hadron jet identification

### 5.1 Data and Monte Carlo samples

The tagging technique presented in this thesis relies on Monte Carlo predictions for the signal (single  $b$ ) or background (merged  $b$ ) hypotheses. The accuracy of the simulation is validated with data by comparing the distributions of the different variables explored.

Samples of jet events from proton-proton collision processes are simulated with PYTHIA8 [27] event generator using a  $2 \rightarrow 2$  matrix element at leading order in the strong coupling to model the hard subprocess, and  $p_T$ -ordered parton showers are utilized to model additional radiation in the leading-logarithmic approximation [28]. Multiple parton interactions [29], as well as fragmentation and hadronisation based on the Lund string model [30] are also simulated. The ATLAS MC11 tune of the soft model parameters was used [31]. In order to have sufficient statistics over the entire  $p_T$  spectrum, eight samples were generated with different thresholds of the hard-scattering

partonic transverse momentum  $\hat{p}_T$ . Events from different samples were mixed taking into account their respective production cross sections.

The GEANT4 [32] software toolkit within the ATLAS simulation framework [33] propagates the generated particles through the ATLAS detector and simulates their interactions with the detector material. The energy deposited by particles in the active detector material is converted into detector signals in the same format as the detector read-out. Finally the Monte Carlo generated events are processed through the trigger simulation package of the experiment, and are reconstructed and analyzed with the same software as for the real data. The simulated data sample used for the analysis gives an accurate description of the pile-up content and detector conditions for the full 2011 data-taking period.

The data samples employed correspond to proton-proton collisions at  $\sqrt{s} = 7$  TeV delivered by the LHC and recorded by ATLAS between May and November 2011, with the LHC running with 50 ns bunch spacing, and bunches organized in bunch trains. Only data collected during stable beam periods in which all sub-detectors were fully operational are used. After the application of the data quality selection, the surviving data corresponds to an integrated luminosity of  $4.7 \text{ fb}^{-1}$ . The LHC performance steadily improved during 2011. In particular the average number of minimum-bias pile-up events, originating from collisions of additional protons in the same bunch as the signal collision, grew from 3 to 20. This fact will be of importance when discussing the selection of discriminating variables.

For the study of systematic effects and for result comparison, other Monte Carlo samples were utilised. Results were produced with the `HERWIG++` generator [8] and with `PYTHIA8` using the Perugia tune [34]. The former is based on the event generator `HERWIG`, but redesigned in the `C++` pro-

gramming language. The generator contains a few modlleing improvements. It also uses angular-ordered parton showers, but with an updated evolution variable and a better phase space treatment. Hadronisation is perfromed using the cluster model. The underlying event and soft inclusive interactions are described using a hard and soft multiple partonic interactions model [35]. The Perugia tune is an independent tune of PYTHIA with increased final state radiation to better reproduce the jet shapes and hadronic event shapes using LEP and TEVATRON data. In addition, paramteres sensitive to the production of particles with strangeness and related to jet fragmentation have been adjusted.

## 5.2 Event and object selection

The event sample for this analysis was collected using a logical OR of single jet triggers which select events with at least one jet with transverse energy above a given threshold at the highest trigger level. The ATLAS Trigger system uses three consecutive trigger levels. At the hardware Level 1 and local software Level 2, cluster-based jet triggers are used to select events. The Level 3, the so-called Event Filter, runs the offline anti- $k_t$  jet finding algorithm with  $R = 0.4$  on topological clusters over the complete calorimeter. At this stage, the transverse energy thresholds, expressed in GeV, are: 20, 30, 40, 55, 75, 100, 135, 180. These triggers reach an efficiency of 99% for events having the leading jet with an offline energy higher than the corresponding trigger thresholds by a factor ranging between 1.5 and 2. The triggers with the lowest  $p_T$  thresholds were prescaled by up to five orders of magnitude, and typically the same jet trigger is prescaled ten times more in the later data taking periods compared to the early ones.

The offline event selection requires at least one primary vertex candidate with 5 or more tracks. All jets, with transverse momentum between 40 and 480 GeV, were required to be in a region with full tracking coverage,  $|\eta_{jet}| < 2.1$ , and they were classified in eight  $p_T$  bins chosen such as to match the jet trigger 99% efficiency thresholds (in GeV): 40, 60, 80, 110, 150, 200, 270, 360. Only jets tagged as  $b$ -jets using the MV1  $b$ -tagging algorithm at the 60% efficiency working point were considered.  $b$ -tagged jets with close-by jets ( $\Delta R < 0.8$ ) with  $p_T$  higher than 7 GeV at electromagnetic scale were not included in the analysis. Unless otherwise indicated, performance plots are shown for one medium  $p_T$  bin (80 to 110 GeV) and one high  $p_T$  bin (200 to 270 GeV).

In the case of MC the reconstructed  $b$ -tagged jets were further classified into single and merged  $b$ -jets based on truth Monte Carlo information. A  $B$ -hadron is considered to be associated to a jet if the  $\Delta R$  distance in  $\eta - \phi$  space between the direction of the hadron and the jet axis is smaller than 0.4. Jets were labeled as merged (single)  $b$ -jets if they contain two (only one)  $B$ -hadron.

It is important to select genuine tracks belonging to jets. Only tracks located within a cone of radius  $\Delta R(jet^{\text{reco}}, \text{track}) \leq 0.4$  around the jet axis were considered. Cuts on  $p_T^{\text{trk}} > 1.0$  GeV and the  $\chi^2$  of the track fit,  $\chi^2/ndf < 3$ , are applied. In addition, tracks are required to have a total of at least seven precision hits (pixel or micro-strip) in order to guarantee at least 3  $z$ -measurements. Tracks are also required to fulfill cuts on the transverse and longitudinal impact parameters at the perigee to ensure that they arise from the primary vertex. As cutting on impact parameter (IP) significance might be detrimental for  $b$ -jets, where large IP values are expected, the relaxed cuts were used,  $|IP_{xy}| < 2$  mm, and  $|IP_z \sin \theta| < 2$  mm, with  $\theta$  being the polar

angle measured with respect to the beam axis.

### 5.3 Preliminary studies in standalone PYTHIA generator

A small set of dijet events was generated using sc pythia. With the help of fastjet [36], anti- $k_t$  jets with distance parameters  $R$  of 0.4 and 0.6 were built using as inputs:

- all stable particles
- charged stable particles
- 0.1  $k_t$  jets, using all particles
- 0.1  $k_t$  jets, using charged particles

The labeling was done in the same procedure as in the Full ATLAS Monte Carlo analysis.

Figures 5.1 to 5.5 show the distributions and correlations of some of the tracking variables, for single and merged  $b$ -jets, in bins of the jet  $p_T$ .

For the single  $b$ -jets,  $\tau_1$ ,  $\tau_2$ , and  $\Delta R$  between the  $k_T$  axes in the jet are all small which is expected for a pencil-like jet. For the  $b\bar{b}$ -jets, these variables are all large, which is typical of a gluon jet. But the correlations are really fascinating in merged  $b$ -jets.  $\tau_1$  and  $\Delta R$  between the  $k_T$  axes are nearly linearly related, which is expected if there are two hard lobes of energy. But  $\tau_2$  is almost independent of  $\Delta R$  between the  $k_T$  axes, meaning that regardless of where the axes are, the energy is uniformly distributed around them.

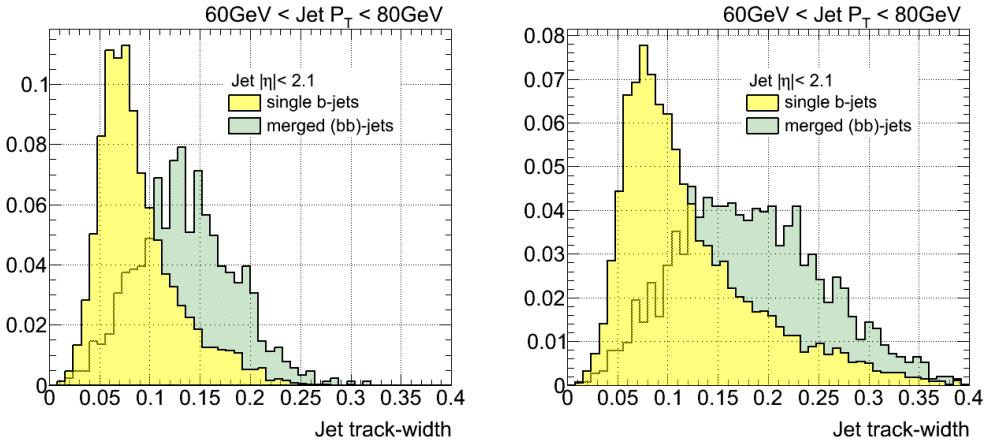


Figure 5.1: Distribution of track-jet width in anti- $k_T$  0.4 (left) and 0.6 (right) jets, for single and merged  $b$ -jets between 60 GeV to 80 GeV. Jets were built using all stable particles in the simulation.

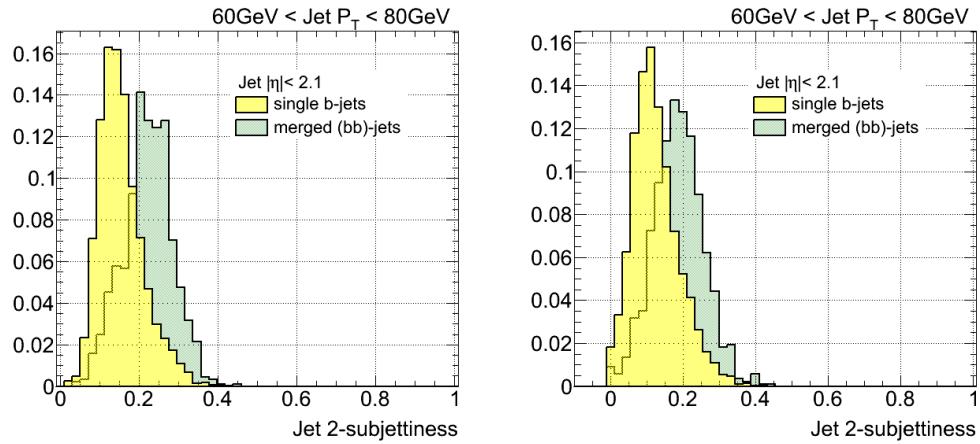


Figure 5.2: Distribution of  $\tau_2$  in anti- $k_T$  0.4 jets, for single and merged  $b$ -jets between 60 GeV to 80 GeV. Jets were built using all stable particles (left) and charged particles only (right).

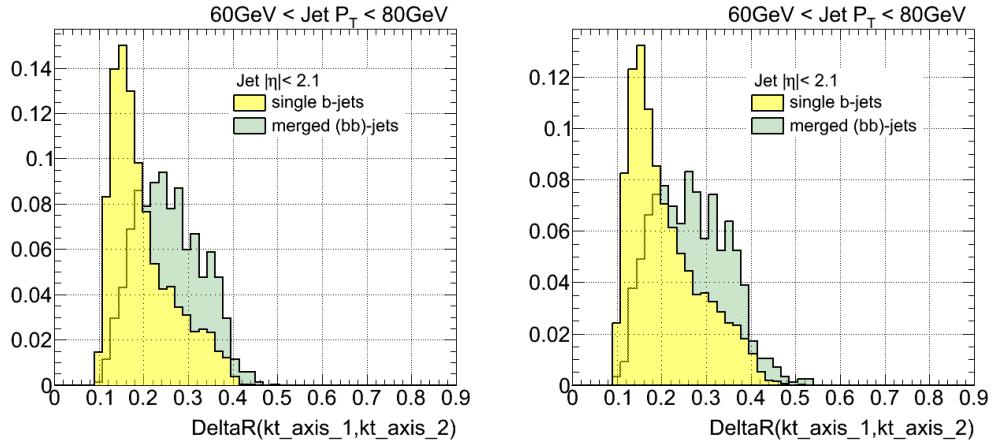


Figure 5.3: Distribution of  $\Delta R$  between the axes of two  $k_T$  subjets in anti- $k_T$  0.4 jets, for single and merged  $b$ -jets between 60 GeV to 80 GeV. Jets were built using 0.1  $k_t$  jets from all stable particles (left) and charged particles only (right).

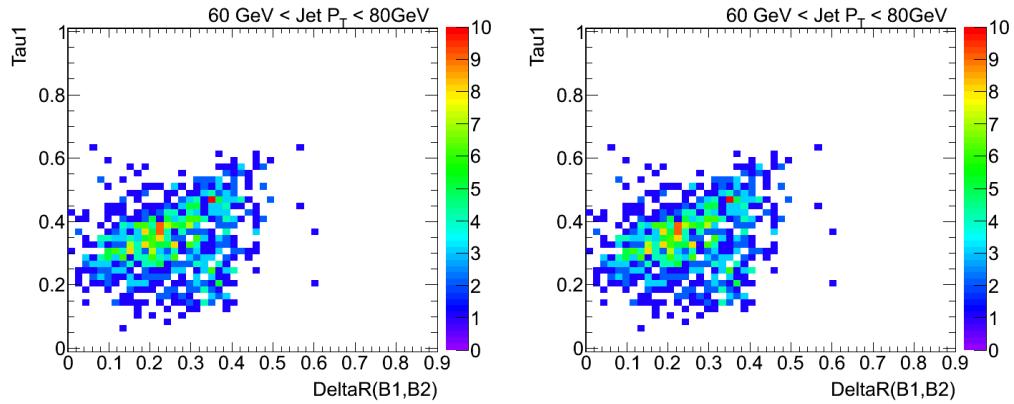


Figure 5.4: Correlation between  $\tau_1$  (left) and  $\tau_2$  (right) and the  $\Delta R$  between the  $B$ -hadrons in merged anti- $k_T$  0.4 jets between 60 GeV to 80 GeV. Jets were built using all stable particles.

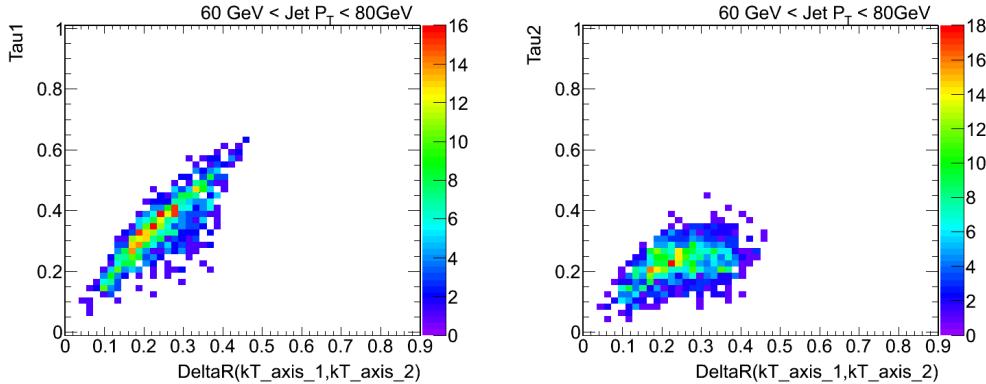


Figure 5.5: Correlation between  $\tau_1$  (left) and  $\tau_2$  (right) and the  $\Delta R$  between the  $k_T$  subjets in merged anti- $k_T$  0.4 jets between 60 GeV to 80 GeV. Jets were built using all stable particles.

## 5.4 Kinematic differences between single and double $B$ -hadron jets

The differences between genuine  $b$ -quark jets and  $b\bar{b}$  jets are expected to arise from the two-subjet (two  $B$ -hadrons) substructure of merged jets. They are thus expected, for the same jet  $p_T$ , to have higher track-multiplicity and be wider than single  $b$ -jets. Based on these characteristics simulated QCD samples of  $b$ -tagged jets were used to study the following properties, discussed in the next paragraphs, built from jet constituents either at calorimeter level (topological clusters) or tracks associated to the jet:

- Jet multiplicity (number of constituents)
- Jet width,  $p_T$  weighted
- Jet Mass
- Nr. of  $k_t$  subjets
- Maximum  $\Delta R$  between pairs of constituents

- $\Delta R$  between 2  $k_t$  subjets within the  $b$ -jet
- $\tau_2$ : 2-subjettiness
- $\tau_2/\tau_1$
- $\Delta R$  of leading constituents
- Eccentricity

### *I. Jet track multiplicity*

This variable is defined as the number of tracks associated to the jet, it is simple to calculate and carries important information of the jet inner structure. Figure 5.6 shows the distribution of the observable for single and merged  $b$ -jets. It was observed that merged  $b$ -jets contain on average around two more tracks than single  $b$ -jets at low jet  $p_T$ , with a larger difference at higher  $p_T$  values. The jet track multiplicity corresponds to tracks with  $p_T$  above 1 GeV, satisfying the quality cuts described in section 5.2. The effect of using a minimum track  $p_T$  of 0.5 GeV was also examined. This was motivated by the fact that it could lead to an improvement in discrimination if it captured more information about the fragmentation process. On the other hand, a lower minimum track  $p_T$  can make the method more sensitive to pile-up with the addition of soft tracks incorrectly associated to the jets. What it was observed is that reducing the  $p_T$  cut only widens the distributions without increasing the separation between single and merged jets.

### *II. Jet width*

The jet width was computed as the  $p_T$  weighted average of the  $\Delta R$  distance between the associated constituent (“*const*”) and the jet axis:

$$Jet\ width = \frac{\sum_{i=1}^N p_T^{const_i} \Delta R(const_i, jet)}{\sum_{i=1}^N p_T^{const_i}} \quad (5.1)$$

where  $N$  is the total number of calorimeter or track constituents.

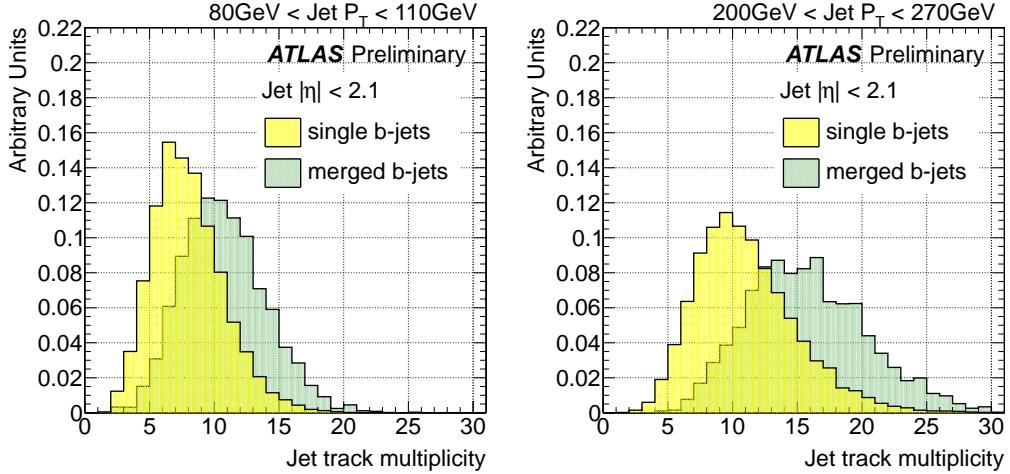


Figure 5.6: Distribution of the track multiplicity in jets for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

Figure 5.7 shows the distribution for the Track-jet width. As expected, merged  $b$ -jets are wider than single  $b$ -jets. In Fig. 5.8 the correlation between the track-jet width and the jet track multiplicity is shown for single and merged  $b$ -jets. These two variables alone provide a good discrimination for tagging  $b\bar{b}$  jets.

The calorimeter jet width ( using topological clusters) gives also good separation. However, this variable is more sensitive to the amount of pile-up in the event than its track-based counterpart. In Fig. 5.9 the distributions of calorimeter width for single and merged  $b$ -jets can be seen for events with low and high Number of Primary Vertices (NPV), in a low  $p_T$  region where the effect of pile-up is more important. In Fig. 5.10 the same distributions are shown for the track-jet width. Calorimeter jet width varies with NPV and due to this behavior the track-based version is more suitable as a more robust discriminator. For similar reasons, the jet topological cluster multiplicity and the jet mass were discarded as discriminating variables.

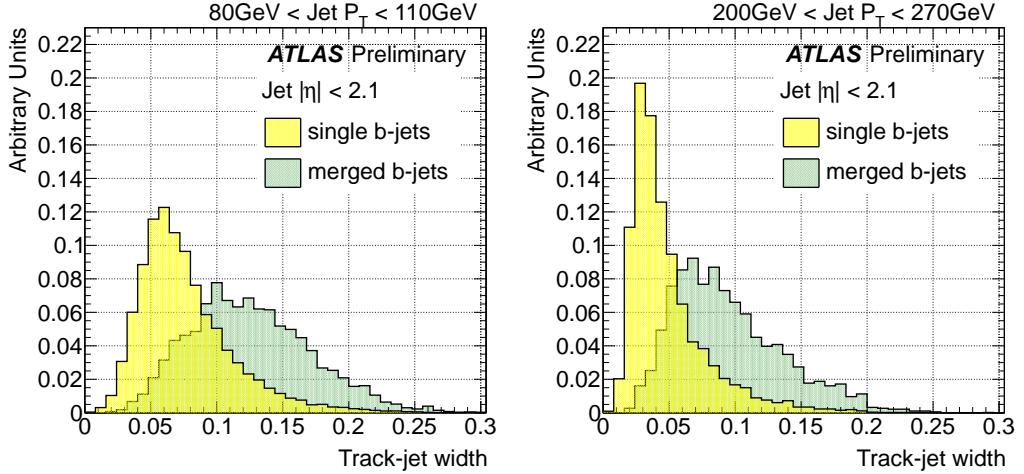


Figure 5.7: Distribution of track-jet width in jets for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

### *III. Maximum $\Delta R$ between track pairs*

Figure 5.11 shows the distribution of the maximum  $\Delta R$  between track pairs in the jets ( $\text{Max}\{\Delta R(\text{trk}, \text{trk})\}$ ). Merged  $b$ -jets show significantly higher values for this variable over a broad range of jet  $p_T$ . The distinct characteristic of this variable is that the separation between single  $b$ -jets and merged does not depend on jet  $p_T$ . In spite of its good discrimination power, we have looked for alternatives to  $\text{Max}\{\Delta R(\text{trk}, \text{trk})\}$  as it is not an infrared safe observable and is sensitive to soft tracks originating from pile-up.

### *IV. $\Delta R$ between the axes of two $k_t$ subjets*

The distribution of the  $\Delta R$  between the axes of the two exclusive  $k_t$  subjets in the jet is shown in Fig. 5.12 for single and merged  $b$ -jets. In order to build this variable the  $k_t$  algorithm [7] is applied to all the tracks associated to

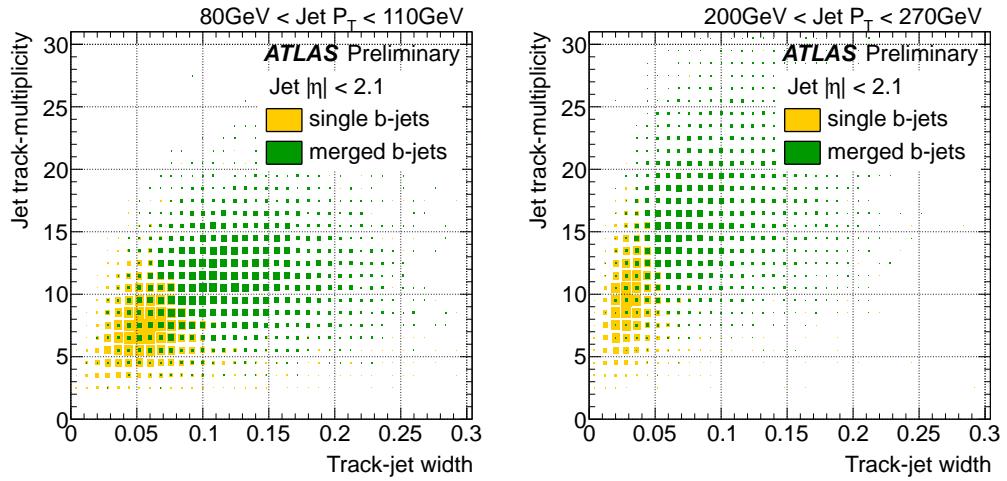


Figure 5.8: Correlation between jet track multiplicity and track-jet width for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

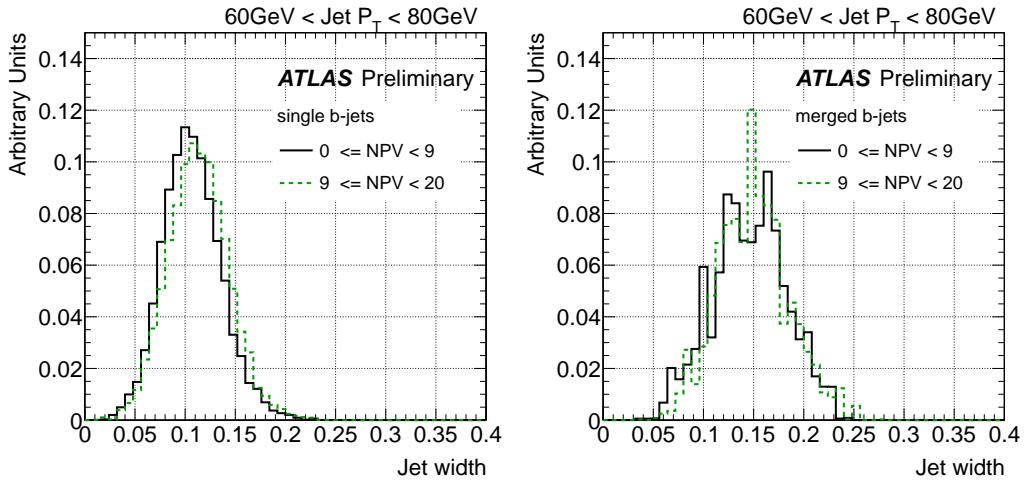


Figure 5.9: Distribution of calorimeter jet width (using topological clusters) for single (left) and merged (right)  $b$ -jets in two bins of Number of Primary Vertices for jets between 60 GeV to 80 GeV.

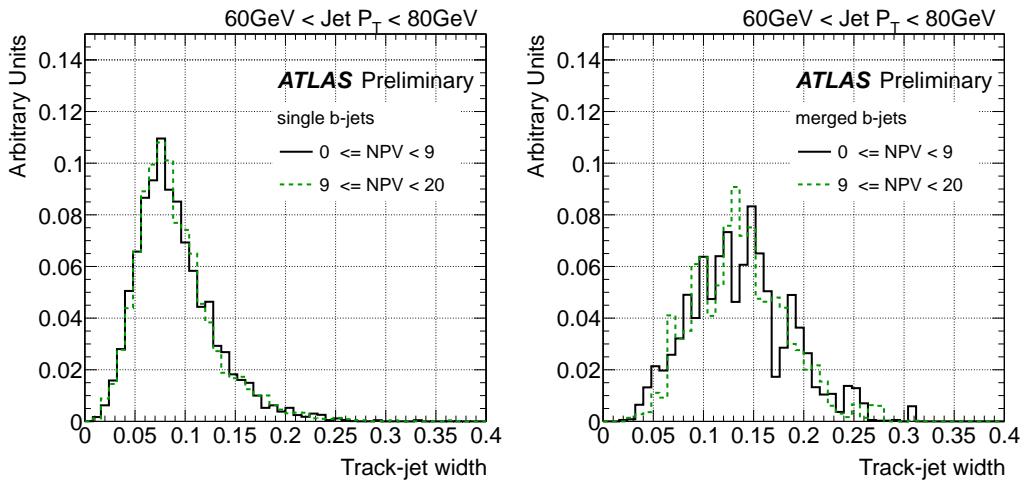


Figure 5.10: Distribution of track-jet width for single (left) and merged (right)  $b$ -jets in two bins of Number of Primary Vertices for jets between 60 GeV to 80 GeV.

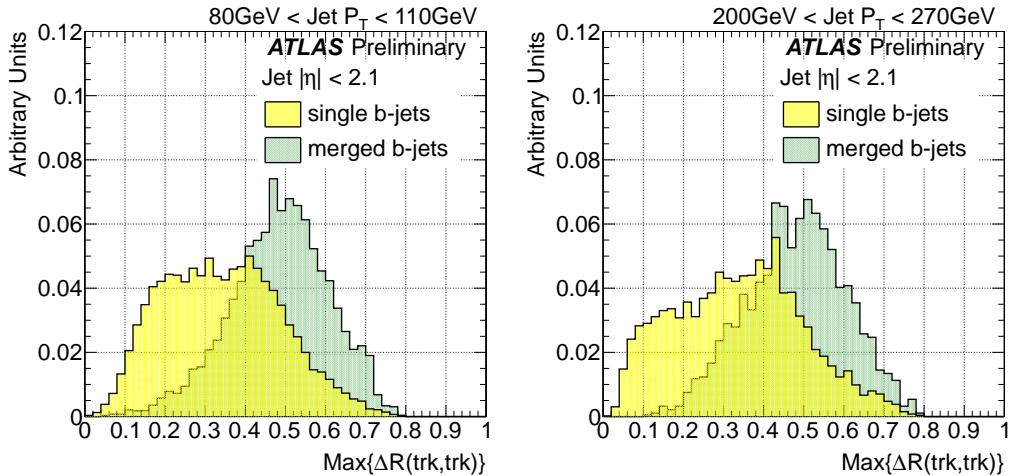


Figure 5.11: Distribution of the maximum  $\Delta R$  between pairs of tracks in jets for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

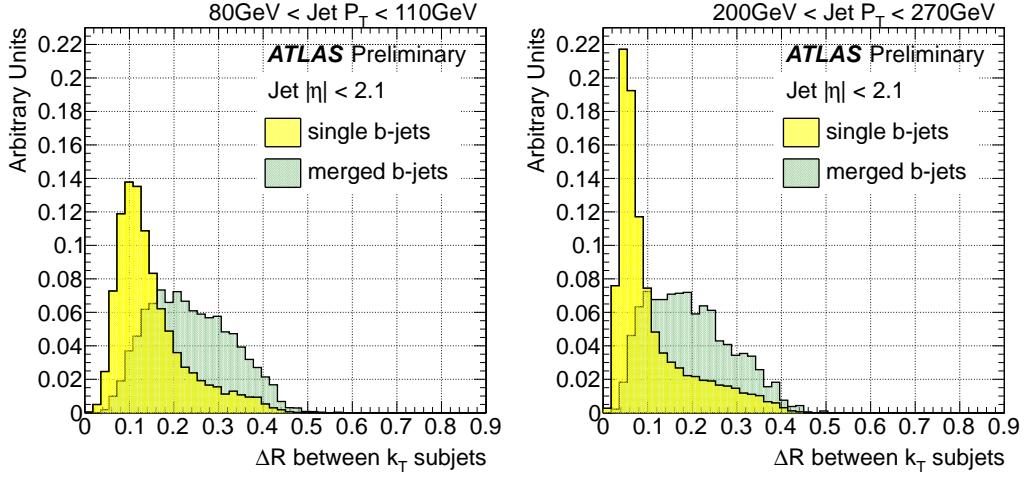


Figure 5.12: Distribution of the  $\Delta R$  between the axes of the two  $k_t$  subjets in the jet for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

the jet using a large  $k_t$  distance parameter to ensure that all of them get clustered. The clustering is stopped once it reaches exactly two jets. We observe that this variable also provides good separation, with the advantage of infrared safeness and insensitivity to pile-up.

### *V. N-subjettiness variables*

$N$ -subjettiness variables, as described in Ref. [37], were originally designed to identify boosted objects, like electroweak bosons and top quarks, decaying into collimated shower of hadrons which a standard jet algorithm would reconstruct as single jets. It is defined as:

$$\tau_N = \frac{1}{\sum_k p_{T,k} R_0} \sum_k p_{T,k} \min\{\Delta R_{S_1,k}, \Delta R_{S_2,k}, \dots, \Delta R_{S_N,k}\} \quad (5.2)$$

where  $R_0$  is the jet radius used in the jet clustering algorithm and the sum runs over the constituents of the jet. To avoid dependence on pile-up we consider the track-based  $n$ -subjettiness, where the sum is over the tracks in the  $b$ -tagged jet.  $\Delta R_{S_j,k}$  is the distance in the rapidity-azimuth plane between the axis of subjet  $j$  and constituent track  $k$ . This jet shape variable quantifies to what degree a jet can be regarded as composed of  $N$  subjets. For instance, a jet with a two pronged structure, with all tracks clustered along two directions, is expected to have a smaller  $\tau_2$  value than a jet with tracks uniformly distributed in  $\eta - \phi$  space.

Plots of  $\tau_2$  are shown in Fig. 5.13. In spite of its expected 2-prong substructure, merged  $b$ -jets have higher values of  $\tau_2$  than single  $b$ -jets. The explanation of this behavior can be found in Fig. 5.14, where its correlation with track-jet width ( $\sim \tau_1$ ) is shown for single and merged  $b$ -jets. The two variables are highly correlated and for this reason wider jets have a larger  $\tau_2$ . This suggests to switch from an absolute to a width-normalized  $\tau_2$ . Fig. 5.15 thus shows the distributions of  $\tau_2/\tau_1$ . This ratio is often used but, although as expected somewhat larger values are obtained for single than for merged  $b$ -jets, specially at high  $p_T$ , we decided not to use this variable as it offers only marginal discrimination.

## VI. Jet Mass

Figure 5.16 shows the distribution of the jet mass for single and merged  $b$ -jets.

## VII. Number of $k_t$ subjets

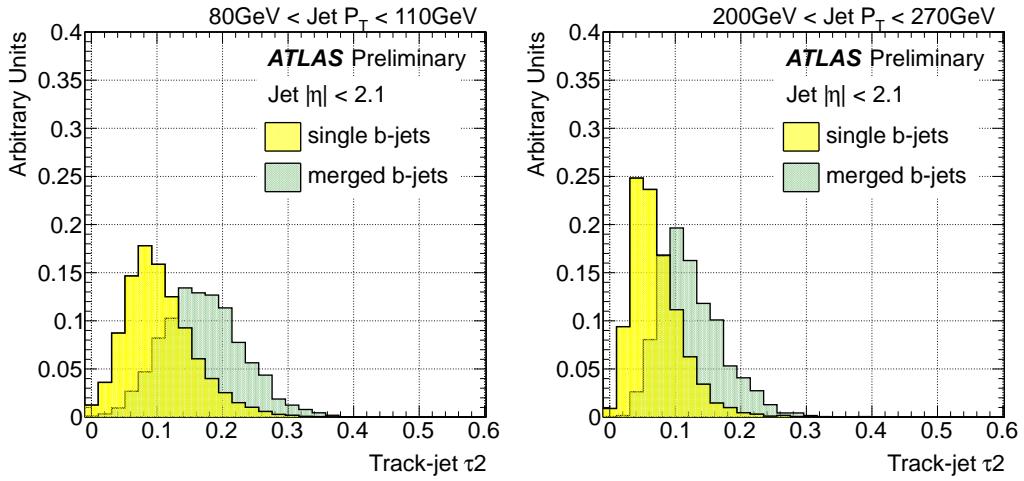


Figure 5.13: Distribution of  $\tau_2$  in jets for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

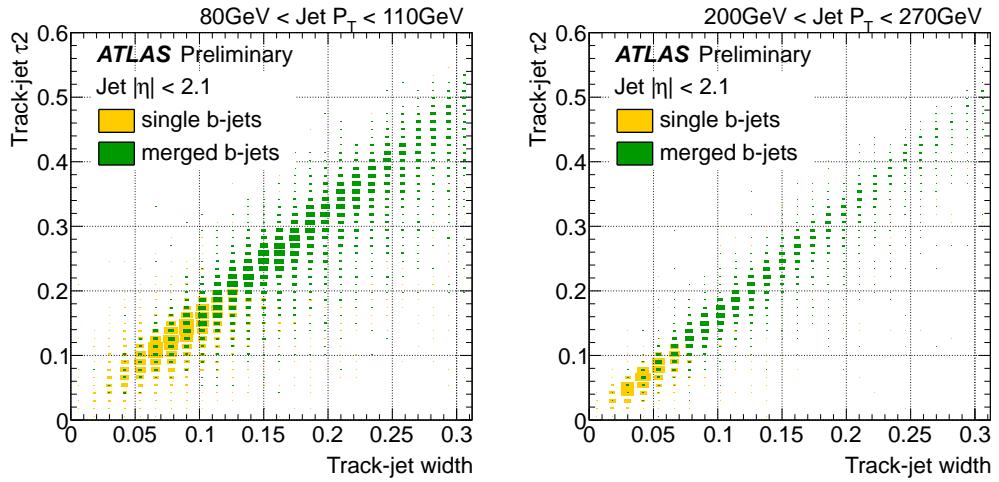


Figure 5.14: Correlation between  $\tau_2$  and track-jet width for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

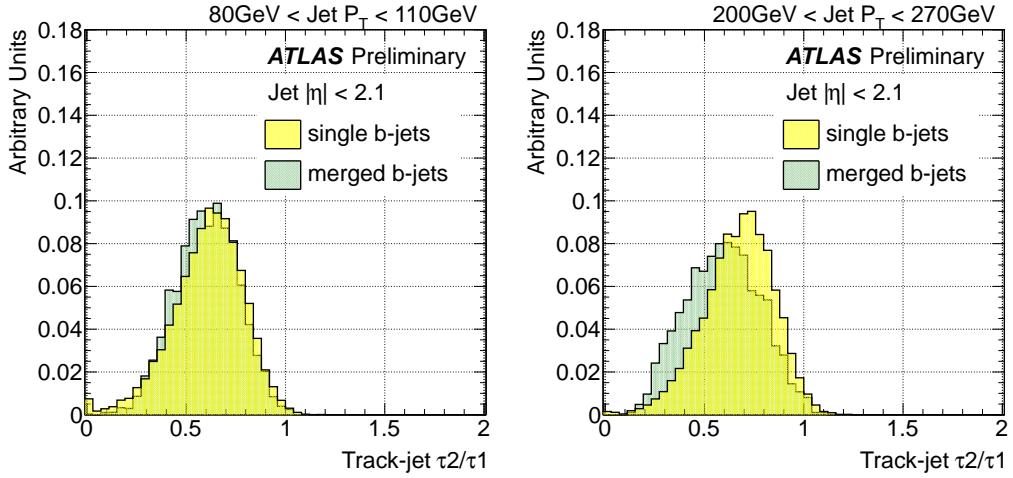


Figure 5.15: Distribution of  $\tau_2/\tau_1$  in jets for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

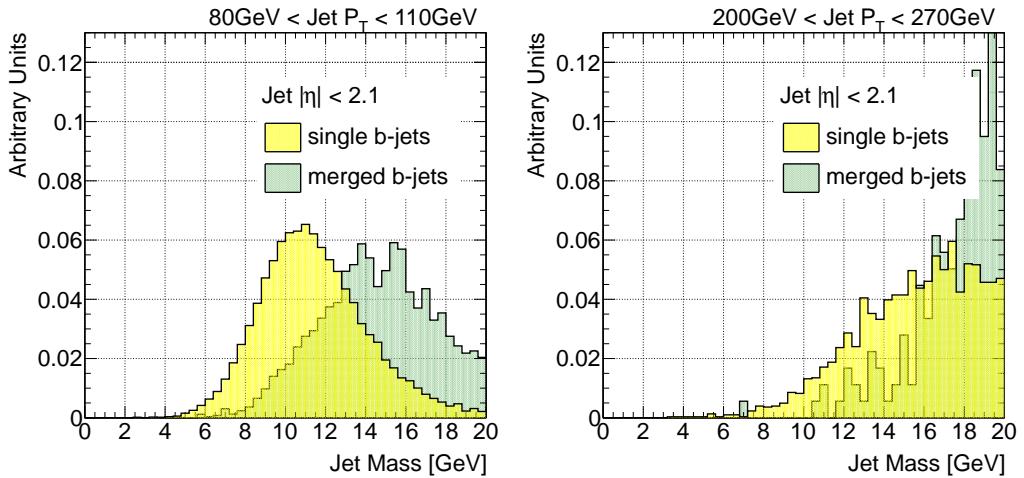


Figure 5.16: Distribution of jet mass in GeV for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

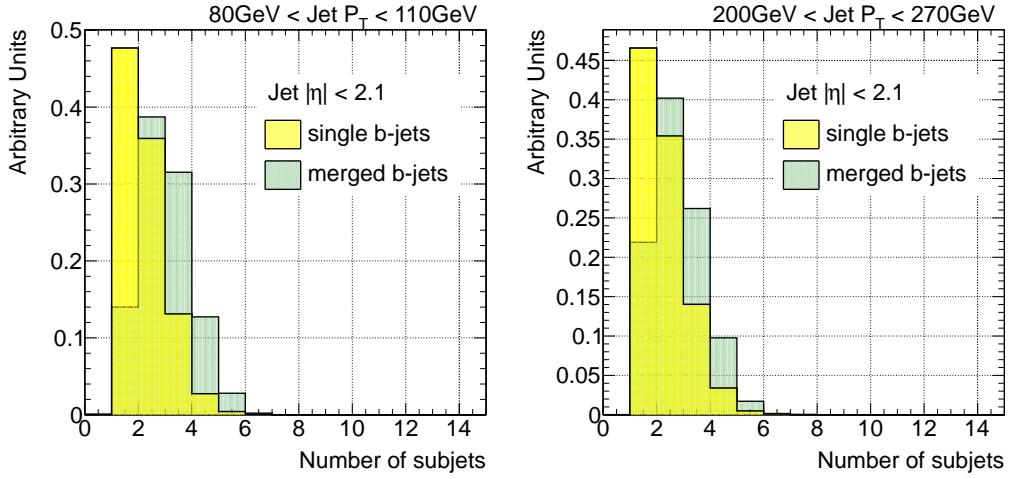


Figure 5.17: Distribution of the number of  $k_t$  sub-track-jets for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

Figure 5.17 shows the distribution of the number of sub-track-jets single and merged  $b$ -jets.

#### *VIII. $\Delta R$ between leading constituents*

Figure 5.18 shows the distribution of the number  $\Delta R$  between leading tracks in the jet for single and merged  $b$ -jets.

#### *IX. Jet eccentricity*

Figure 5.19 shows the distribution of the jet track-eccentricity for single and merged  $b$ -jets.

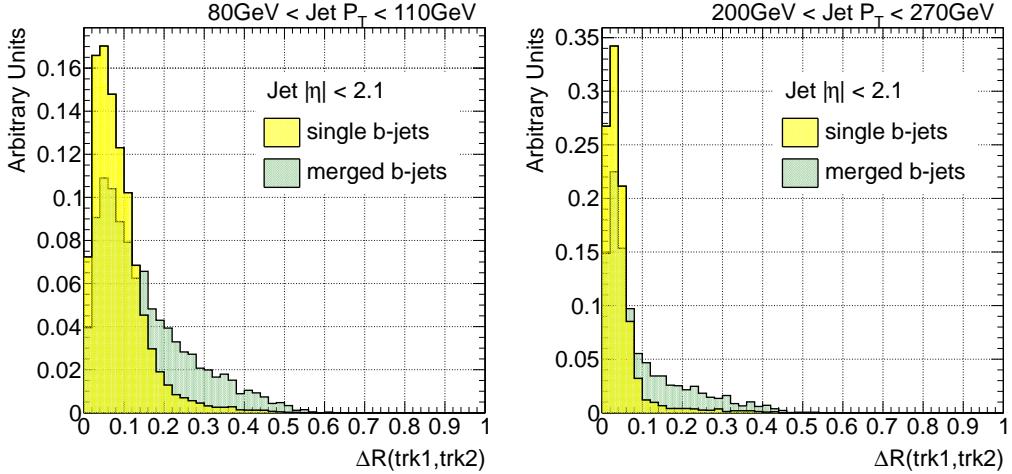


Figure 5.18: Distribution of  $\Delta R$  between leading tracks for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

We also explored the potential improvement of constructing kinematic variables with only displaced tracks, as these are the ones expected to arise from the decay of  $B$ -hadrons. Cuts of 2, 2.5 and 3 on the track transverse impact parameter significance were investigated leading however to no gain in discrimination power.

In Figures 5.20 and 5.21 two examples are shown.

#### 5.4.1 Further studies using “ghost-association” and bigger cone jets

In order to better understand the behavior observed for  $\tau_2$ ,  $\Delta R$  between the axes of  $k_T$  subjets and jet eccentricity in anti- $k_T$  0.4 jets, these variables were studied for other two different scenarios,

- using the active area of jets (with clusters used as input to jet recon-

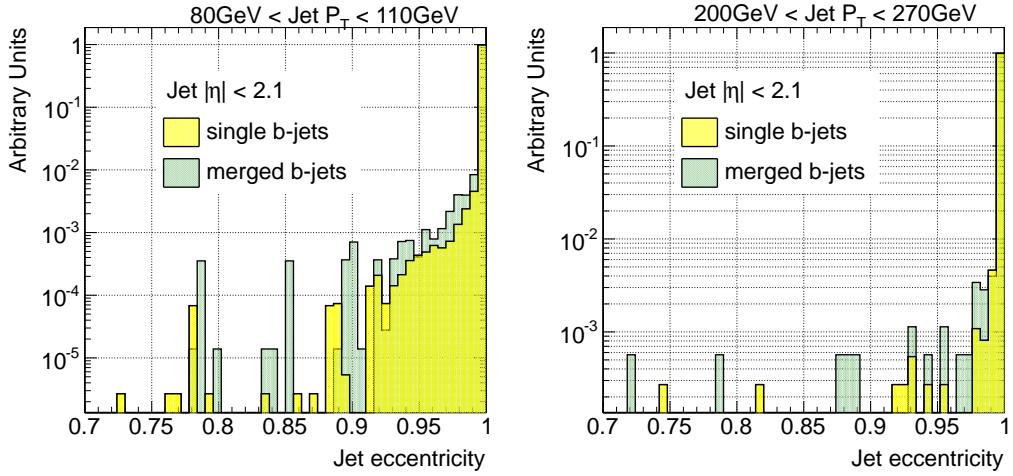


Figure 5.19: Distribution of the jet eccentricity for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

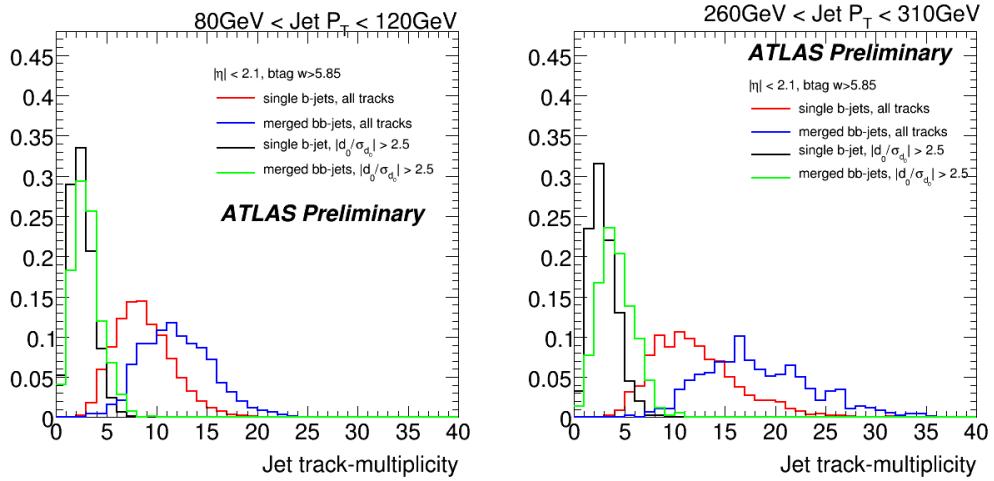


Figure 5.20: Distribution of the jet track multiplicity single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right), for all and displaced tracks only.

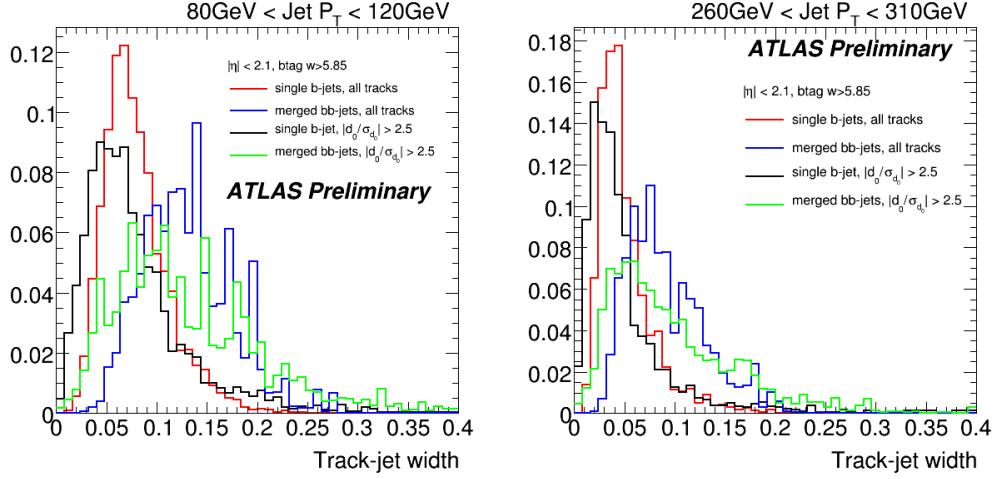


Figure 5.21: Distribution of the track-jet width for single and merged  $b$ -jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right), for all and displaced tracks only.

struction).

- using bigger 0.6 anti- $k_T$  jets

in order to enhance the efficiency to capture the decay products in gluon to  $b\bar{b}$ -jets.

Figures 5.22 to 5.24 show distributions of variables mentioned above for single and merged  $b$ -jets between 80 GeV to 110 GeV.

## 5.5 Validation of the jet variables in data

In order to study the extent to which the simulation reproduces the distributions observed in data for the different variables explored a set of comparison plots is presented. Fig. 5.25 shows the distributions of jet track multiplicity, track-jet width and  $\Delta R$  between the axes of the two  $k_t$  subjets, in two different jet  $p_T$  bins in dijet Monte Carlo and data events collected by AT-

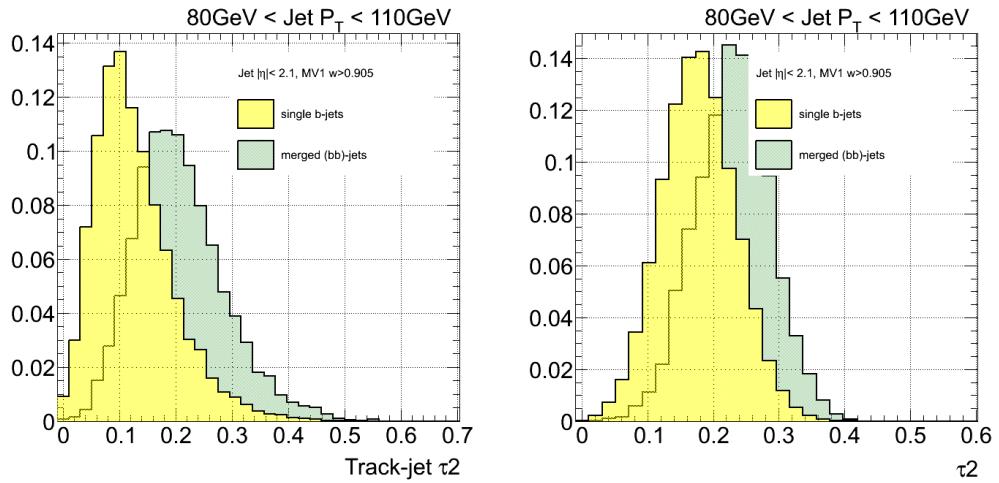


Figure 5.22: Distribution of  $\tau_2$  for single and merged  $b$ -jets between 80 GeV to 110 GeV in anti- $k_T$  0.6 jets using track constituents (left) and anti- $k_T$  0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

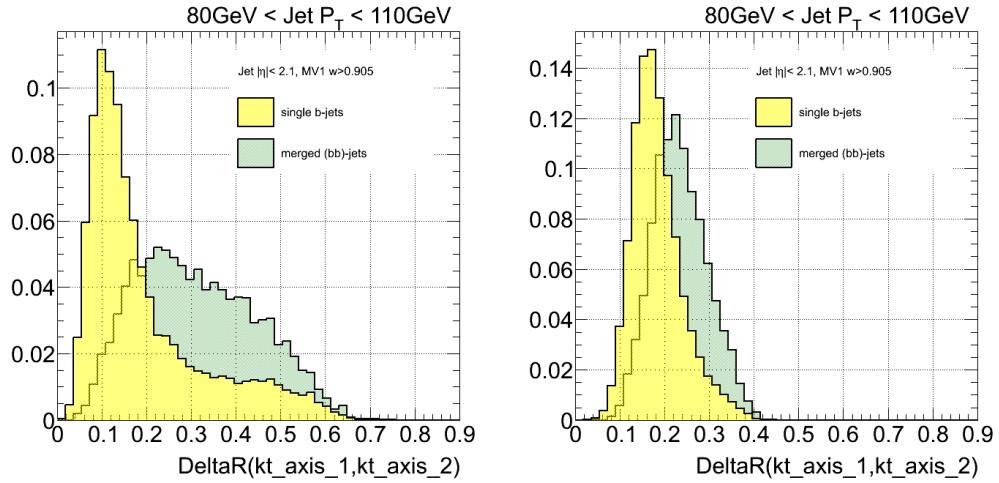


Figure 5.23: Distribution of  $\Delta R$  between  $k_T$  subjets for single and merged  $b$ -jets between 80 GeV to 110 GeV in anti- $k_T$  0.6 jets using track constituents (left) and anti- $k_T$  0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

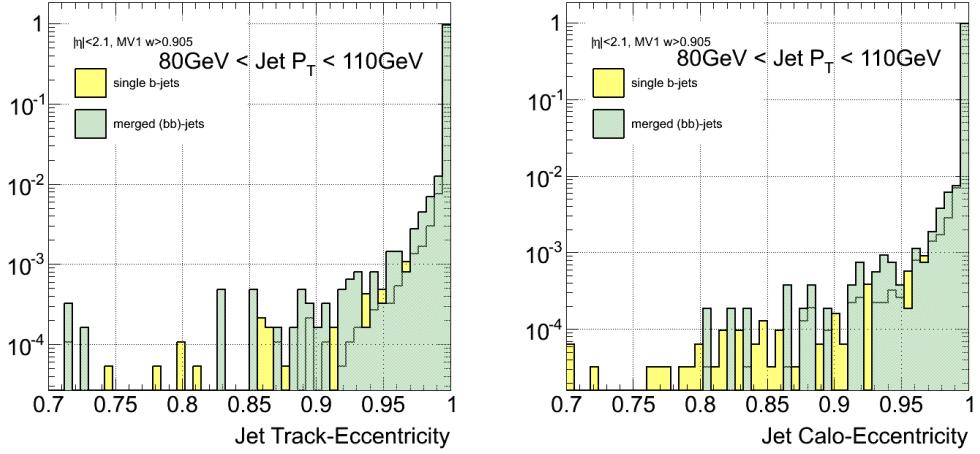


Figure 5.24: Distribution of the jet eccentricity for single and merged  $b$ -jets between 80 GeV to 110 GeV in anti- $k_T$  0.6 jets using track constituents (left) and anti- $k_T$  0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

LAS during 2011. The distributions are normalized to unit area to allow for shape comparisons. There is a good agreement between data and simulation. It should be remarked that the observed agreement is actually not a direct validation of the description in the MC of the relevant variables, but its convolution with the simulated relative fractions of light-,  $c$ -,  $b$ - and  $bb$ -jets in the  $b$ -tagged generated jet sample. To some extent, some level of compensation can take place between these two effects.

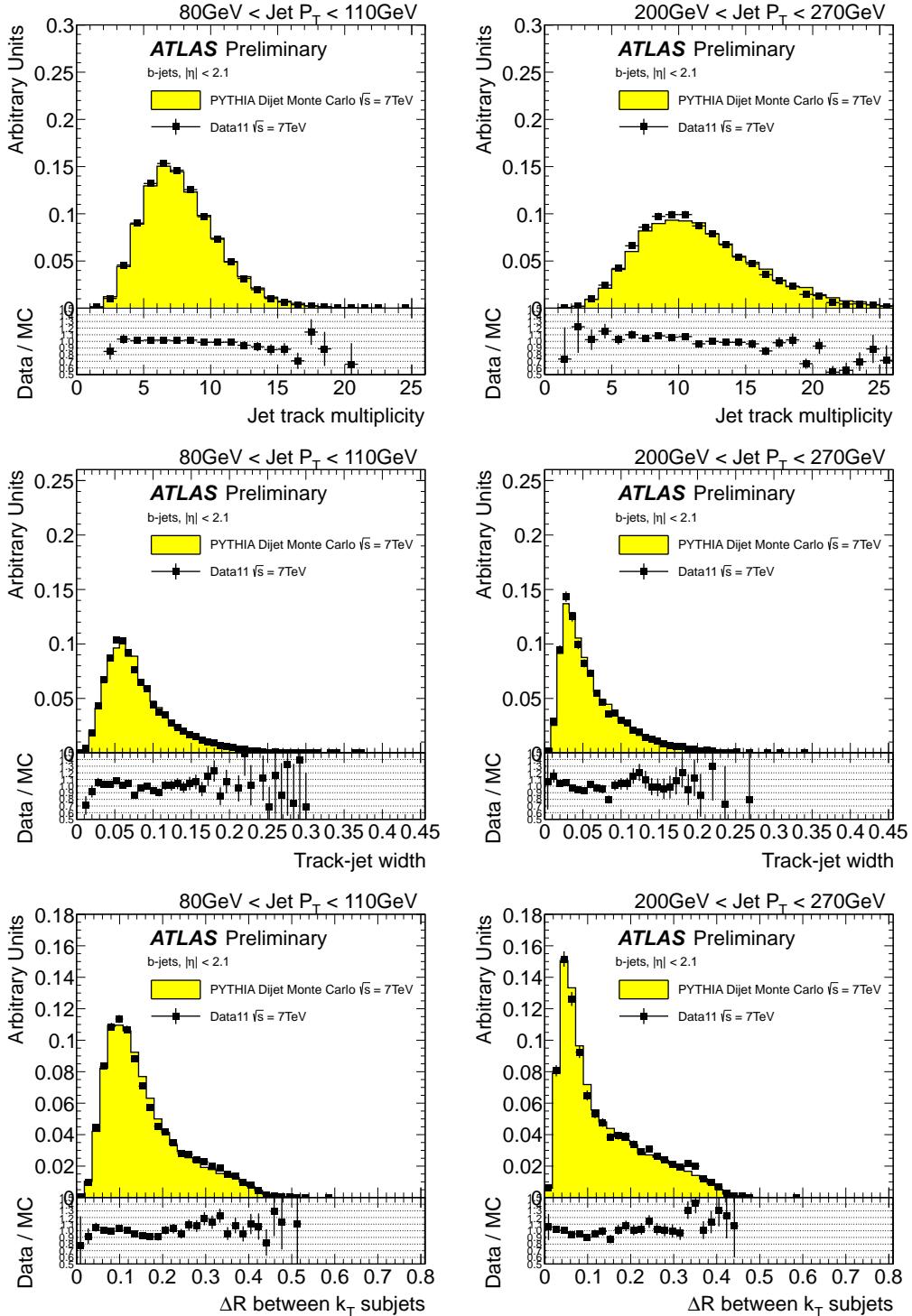


Figure 5.25: Distribution of three tracking variables in 2 different jet  $p_T$  bins, for experimental data collected by ATLAS during 2011 (solid black points), and simulated data (filled histograms). The ratio data over simulation is shown at the bottom of each plot.

# Chapter 6

## Multivariate Analysis

### 6.1 The multivariate classifiers

The following multivariate methods were explored:

- Likelihood ratio estimators
- Neural Networks (NN)
- Boosted decision Trees (BDTs)

And different trainings were tested:

- Inclusive, with  $p_T$ -weighting
- In bins of jet  $p_T$

Signal and background jets were not weighted by the dijet samples cross-sections to allow the contribution of subleading lower  $p_T$  jets from high  $p_T$  events, and thus increase the statistics of merged jets in the low  $p_T$  bins.

Figure 6.1 and 6.2 show distributions of the MVA outputs in different bins of jet  $p_T$  for the two proposed trainings. In figures 6.3 and 6.4 a comparison of the performance of all methods, for inclusive and “in-bins”, training is illustrated.

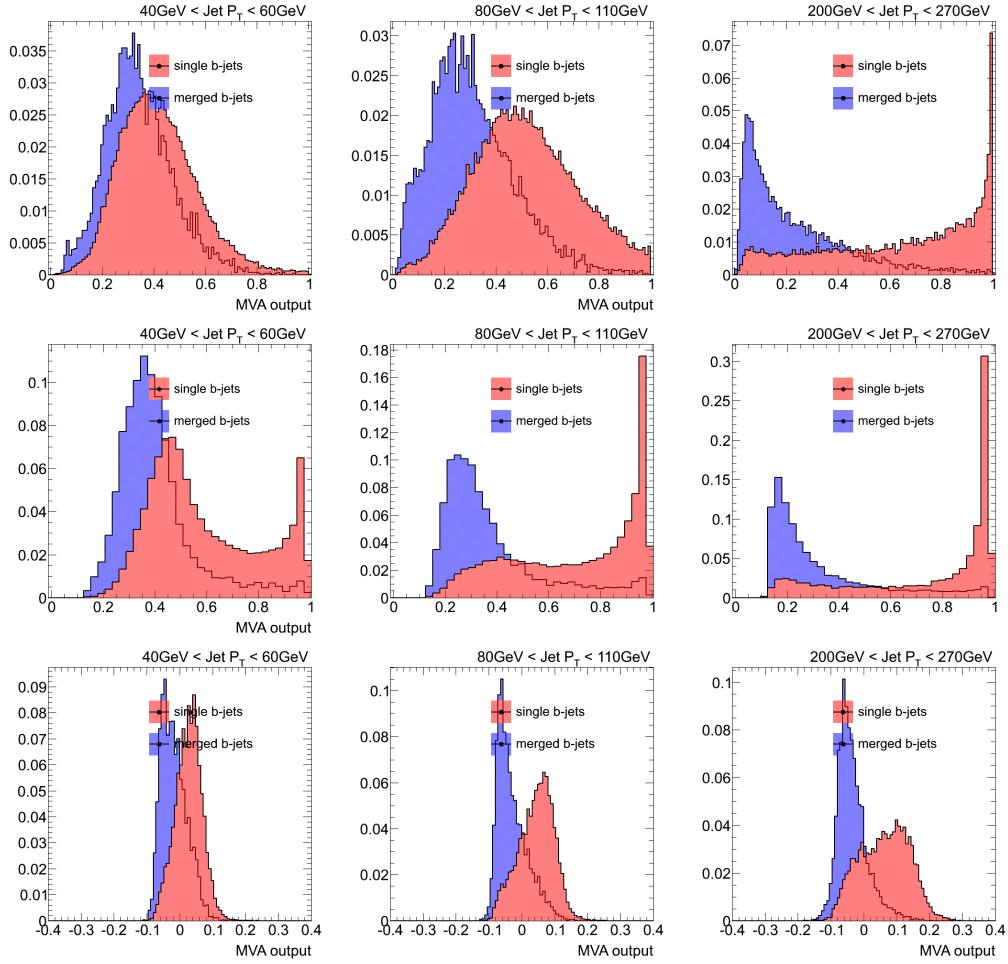


Figure 6.1: Distribution of the MVA discriminant outputs, for inclusive training, in single and merged  $b$ -jets, for low, medium and high jet  $p_T$ .

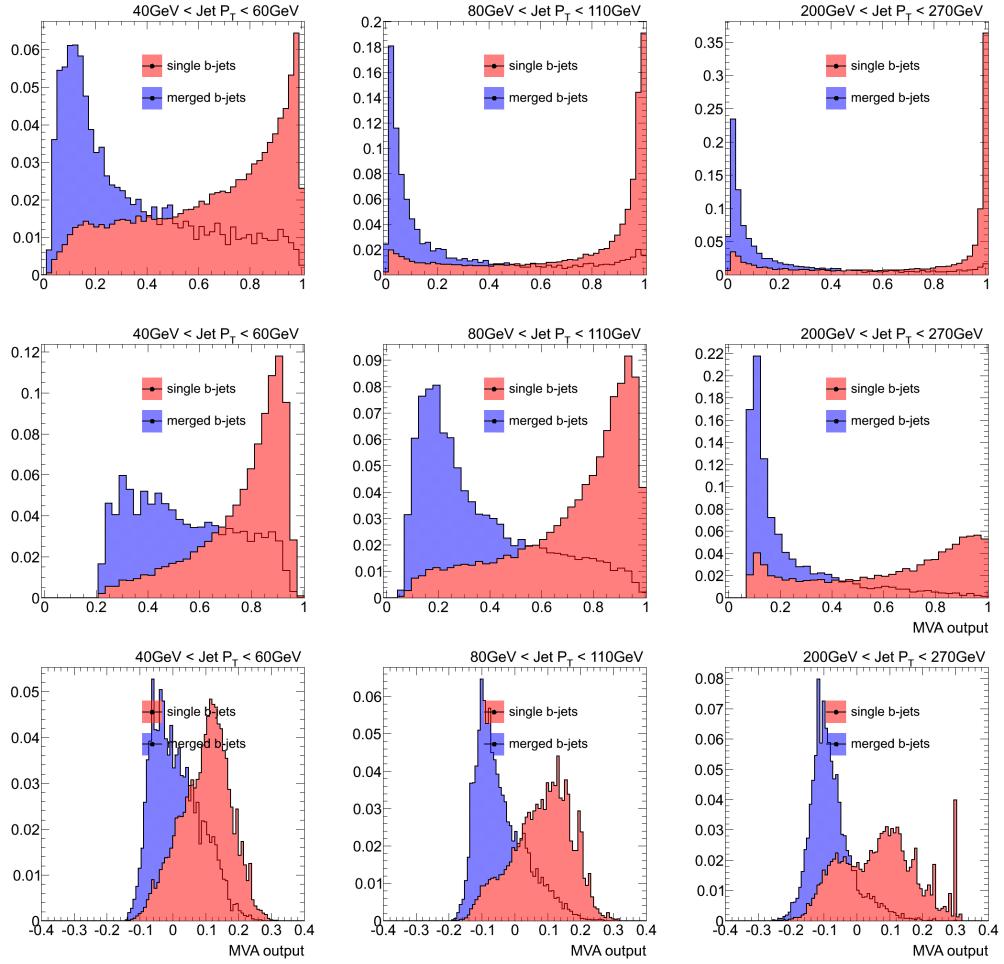


Figure 6.2: Distribution of the MVA discriminant outputs, for training in bins of jet  $p_T$ , in single and merged  $b$ -jets, for low, medium and high jet  $p_T$ .

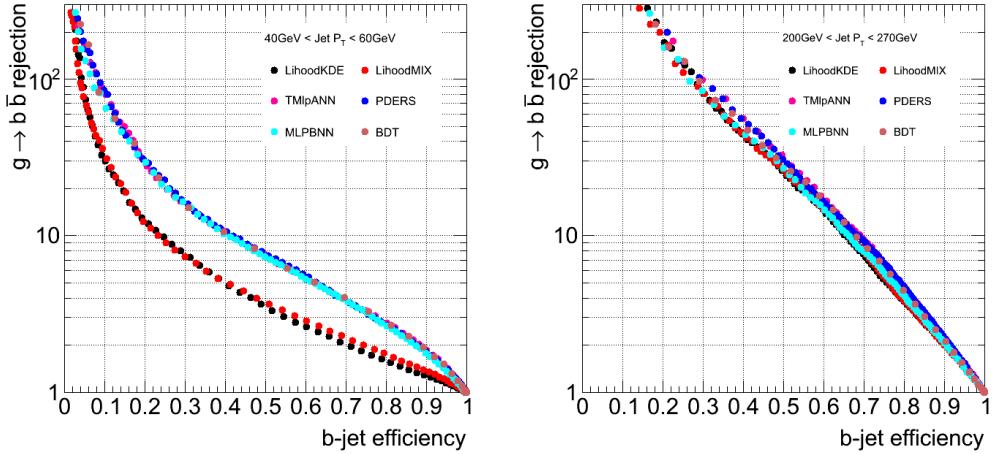


Figure 6.3: Distribution of the MVA discriminant performance for inclusive training, in single and merged  $b$ -jets, for low and high jet  $p_T$ .

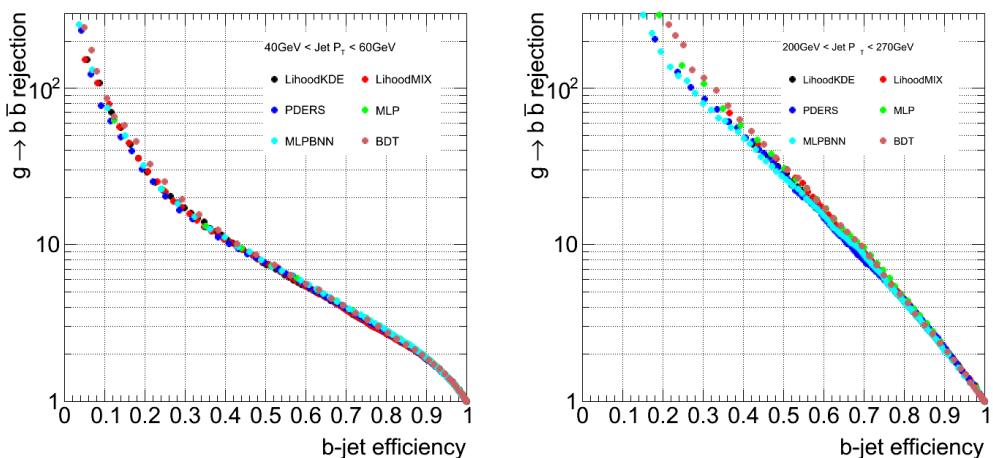


Figure 6.4: Distribution of the MVA discriminant performance for training in bins of jet  $p_T$ , in single and merged  $b$ -jets, for low and high jet  $p_T$ .

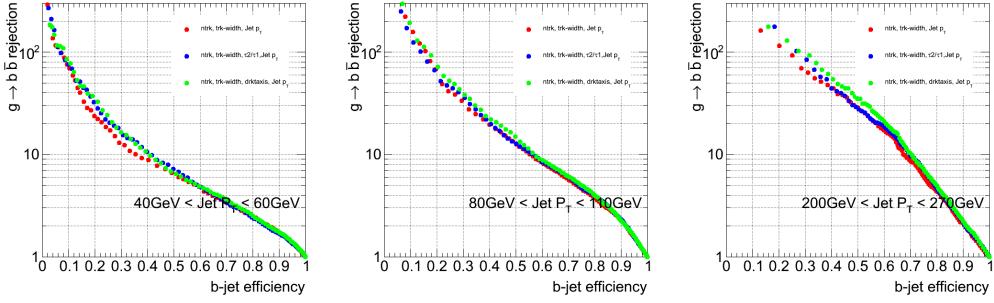


Figure 6.5: Distribution of the MVA discriminant performance for three sets of input variables, in single and merged  $b$ -jets, for low, medium and high jet  $p_T$ .

## 6.2 The input variables

Different groups of input variables were tested. Figure ?? shows the performance for three sets of variables for MVA classifier.

## 6.3 $g \rightarrow b\bar{b}$ likelihood training and performance

A discriminant between single  $b$ -jets and merged  $b$ -jets was built by training a simple likelihood estimator in the context of the Toolkit for Multivariate Data Analysis, TMVA [38].

A sub-set of the dijet Monte Carlo sample was used for training. After the event and jet selections were performed, the  $b$ -tagged jets with  $|\eta| < 2.1$  were classified as signal (single  $b$ -jets) or background (merged  $b$ ). The likelihood training was done in bins of calorimeter jet  $p_T$ . Signal and background jets were not weighted by the dijet samples cross-sections to allow the contribution of subleading lower  $p_T$  jets from high  $p_T$  events, and thus increase the statistics of merged jets in the low  $p_T$  bins. For the evaluation of the method the same procedure was followed.

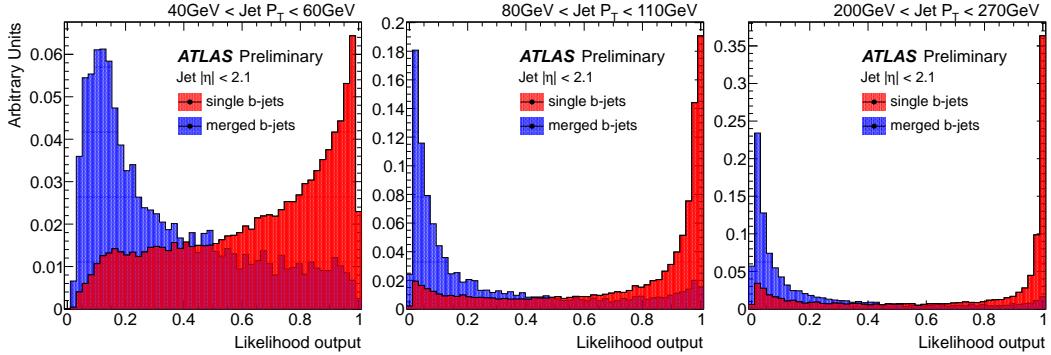


Figure 6.6: Distribution of the  $g \rightarrow b\bar{b}$  likelihood output for single and merged  $b$ -jets for low, medium and high  $p_T$  jets.

Several combinations of the tracking and jet shape variables studied in the previous section were tested as input variables. We found that the following three offer the best performance:

1. Jet track multiplicity
2. Track-jet width
3.  $\Delta R$  between the axes of 2  $k_t$  subjets within the jet

A requirement of at least two matching tracks was imposed to all  $b$ -tagged jets in order to build the third variable listed. This cut was applied in both training and testing samples.

The distribution of the likelihood output for single and merged  $b$ -jets is shown in Fig. 6.6 for low, medium and high transverse momentum jets.

The performance of the  $g \rightarrow b\bar{b}$  tagger in the simulation can be displayed in a plot of rejection ( $1/\epsilon_{bkg}$ ) of merged  $b$ -jets as a function of single  $b$ -jet efficiency, where  $\epsilon_{bkg}$  is the probability that a  $b\bar{b}$ -jet passes the tagger. This is shown in Fig. 6.7 for the eight bins of jet  $p_T$  mentioned in section 5.2. The performance improves with  $p_T$ :

- $p_T > 40$  GeV: rejection above 8 at 50% eff.

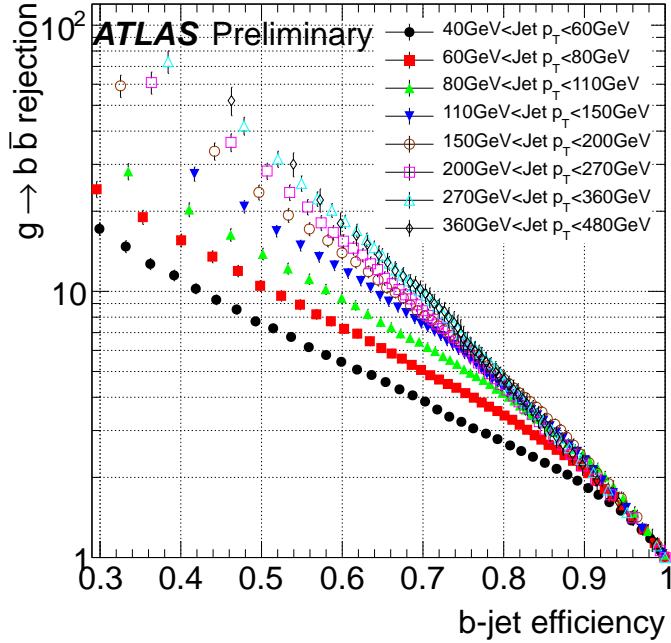


Figure 6.7: Rejection of  $g \rightarrow b\bar{b}$  merged  $b$ -jets as a function of  $b$ -jet efficiency for dijet events in 8 jet  $p_T$  bins.

- $p_T > 60$  GeV: rejection above 10 at 50% eff.
- $p_T > 200$  GeV: rejection above 30 at 50% eff.

The likelihood was trained with jets that had been first tagged by the MV1 algorithm. In order to use the  $g \rightarrow b\bar{b}$  classifier for jets tagged by another tagger a new training is required.

The rejection of merged jets attained as a function of  $p_T$  for the 50% and 60% efficiency working points are summarized in Table 6.1, together with their relative statistical error. These are propagated from the Poisson fluctuations of the number of events in the merged and single  $b\bar{b}$  distributions. The error is slightly lower for the 60% efficiency working point because a higher efficiency allows for a greater number of Monte Carlo events to measure

the performance.

Jet $p_T$ (GeV )	single $b$ -jet efficiency 50%		single $b$ -jet efficiency 60%	
	Rejection	stat.err.	Rejection	stat.err.
40 - 60	8	4%	5	3%
60 - 80	10	4%	7	4%
80 - 110	14	5%	9	4%
110 - 150	19	5%	12	4%
150 - 200	23	5%	14	5%
200 - 270	30	7%	16	6%
270 - 360	36	7%	19	6%
360 - 480	41	8%	18	8%

Table 6.1: The merged  $b$ -jet rejection for the 50% and 60% efficiency working points in bins of  $p_T$ .

## 6.4 Systematic uncertainties

The development, training and performance determination of the tagger is based on simulated events. Although the agreement between simulation and data explored in section ?? is a necessary validation condition, it is also important to investigate how the tagger performance depends on systematics relevant in the data. In particular we have considered:

- presence of additional interactions (pile-up)
- uncertainty in the  $b$ -jet tagging efficiency
- uncertainty in the track reconstruction efficiency

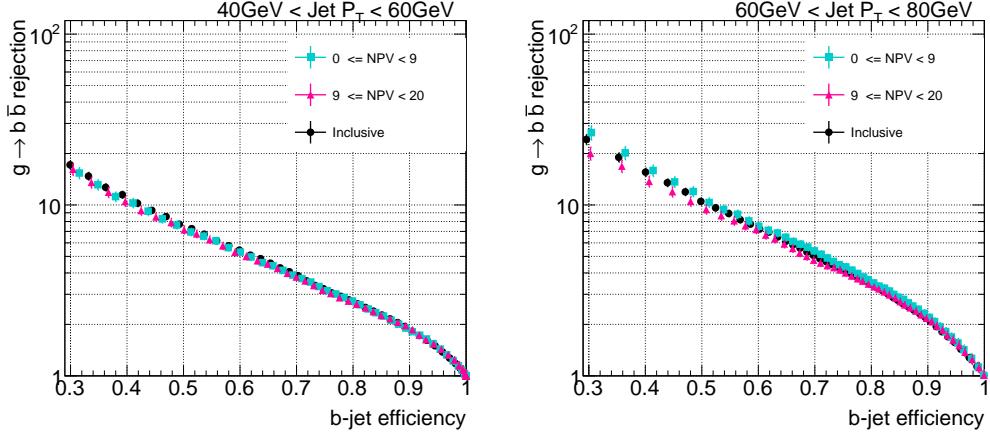


Figure 6.8: Rejection of  $g \rightarrow b\bar{b}$  merged b-jets as a function of  $b$ -jet efficiency in bins of  $N_{\text{vtx}}$  for two low jet  $p_T$  bins.

- uncertainty in the track transverse momentum resolution
- uncertainty in the jet transverse momentum resolution

### *I. Pile-up*

The size of this effect was studied by comparing the performance of the likelihood discriminant with  $b$ -jets in events with small (1-9) and large (9-20) number of primary vertices. The comparison of the performance in these two sub-samples can be seen in Fig. 6.8. As expected from the use of tracking (as opposed to calorimeter) variables no significant dependence with pile-up is observed within statistics. Of the 16 determinations (2 working points with 8  $p_T$  bins each) of performance differences between high and low number of primary vertices events, it is observed that 6 of them are positive and 10 negative, with a global mean of 0.3%. We conclude that the effect is negligible compared to other source of uncertainties.

### *II. $b$ -tagging efficiency*

The performance of heavy-flavor tagging in Monte Carlo events is calibrated

to experimental data by means of the scale factors (SFs) measured by the *b*-tagging group. Such a measurement carries a systematic uncertainty, and in order to estimate its effect a conservative approach is followed: the SFs are varied in all the  $p_T$  bins simultaneously by one standard deviation both in the up and down directions. The result of this procedure for the distribution of two of the tracking variables used in our discriminant is illustrated in Fig. 6.9.

The effect of the *b*-tagging calibration uncertainty on the likelihood performance is  $< 1\%$ , negligible with respect to the statistical uncertainty as it can be seen in Fig. 6.10. This was indeed expected. The scale factors depend on the true flavor of the jet and on its  $p_T$ , but these are basically constant in the performance determination, which is based on single flavor (true *b*-) jets classified in  $p_T$ -bins.

### *III. Track reconstruction efficiency*

This uncertainty arises from the limit in the understanding of the material layout of the Inner Detector. To test its impact a fraction of tracks determined from the track efficiency uncertainty was randomly removed following the method in Ref. [39].

The tracking efficiency systematics are given in bins of track  $\eta$ . For tracks with  $p_T^{\text{track}} > 500$  MeV the uncertainties are independent of  $p_T$ : 2% for  $|\eta^{\text{track}}| < 1.3$ , 3% for  $1.3 < |\eta^{\text{track}}| < 1.9$ , 4% for  $1.9 < |\eta^{\text{track}}| < 2.1$ , 4% for  $2.1 < |\eta^{\text{track}}| < 2.3$  and 7% for  $2.3 < |\eta^{\text{track}}| < 2.5$  [18]. All numbers are relative to the corresponding tracking efficiencies.

The tracking variables were re-calculated and the performance of the nominal likelihood was evaluated in the new sample with worse tracking efficiency. The rejection-efficiency plots, shown in Fig. 6.11, show a small degradation

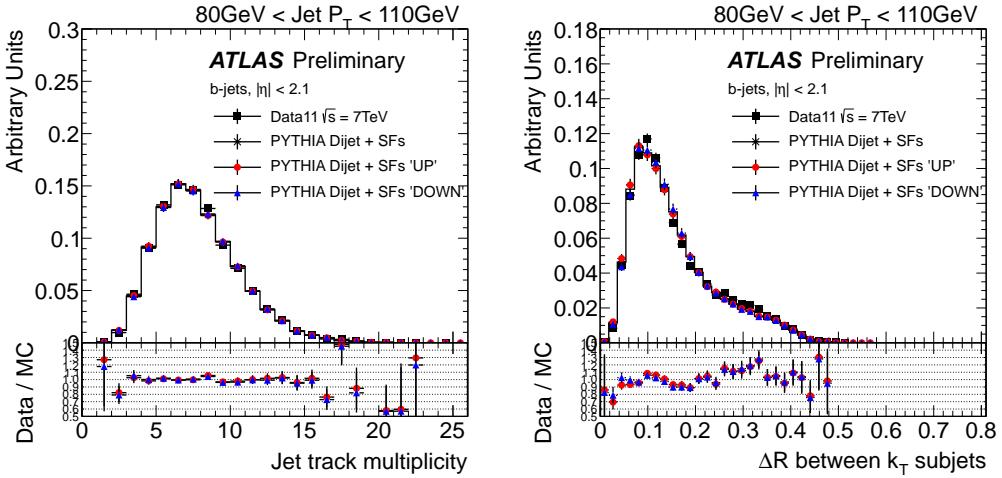


Figure 6.9: The effect of a variation in the  $b$ -tagging Scale Factors on the tracking variables distributions. Scale Factors were varied up (down) by 1-sigma to evaluate the systematic uncertainty from this source. The ratio data over MC is shown for MC PYTHIA with SFs varied up (circles) and down (triangles).

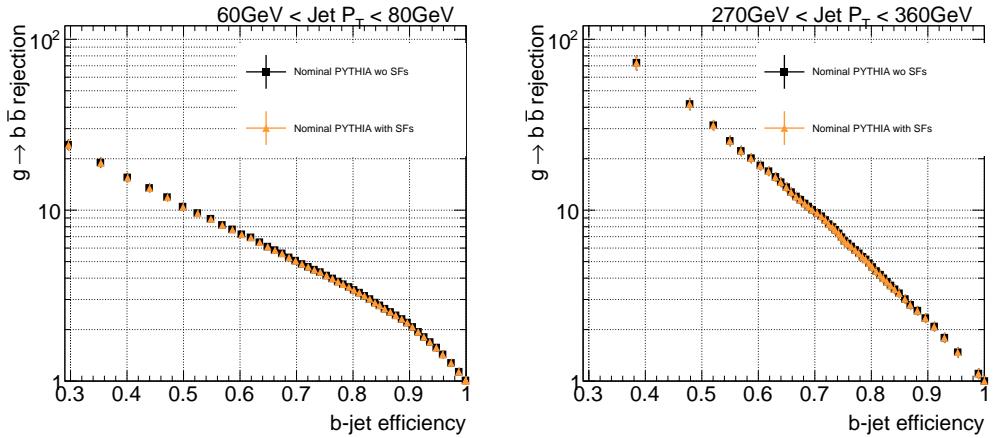


Figure 6.10: Rejection of  $g \rightarrow b\bar{b}$  merged b-jets as a function of  $b$ -jet efficiency with and without scale factors.

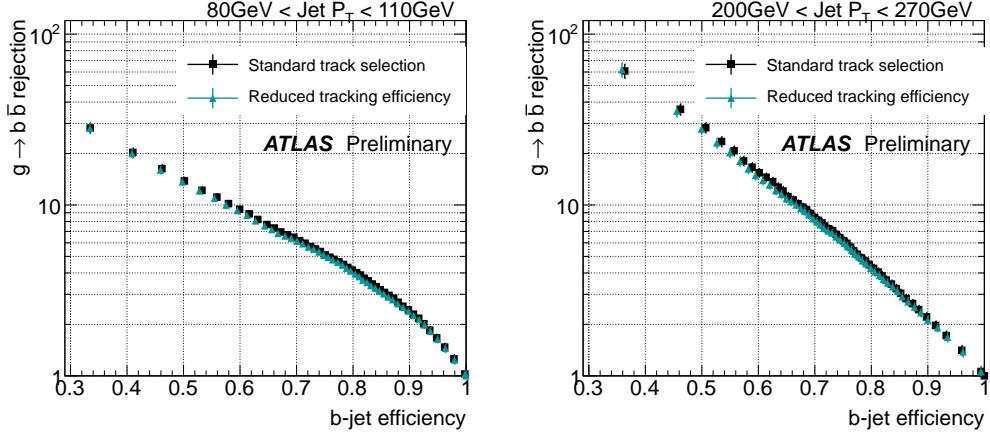


Figure 6.11: Rejection of  $g \rightarrow b\bar{b}$  merged  $b$ -jets as a function of  $b$ -jet efficiency showing shift in likelihood performance caused by a reduction in the tracking efficiency .

of the performance which is comparable to the statistical uncertainty. The effect is however systematically present over all 16  $p_T$  bin/working points, without a clear  $p_T$  dependence. We have thus taken the average over  $p_T$ , and obtained a global systematic uncertainty of 4% both for the 50% and 60% efficiency working points.

#### *IV. Track momentum resolution*

The knowledge of the track momentum resolution is limited by the precision both in the material description of the Inner Detector and in the mapping of the magnetic field. Its uncertainty propagates to the kinematic variables used in the  $g \rightarrow b\bar{b}$  tagger. In order to study this effect, track momenta are oversmeared according to the measured resolution uncertainties before computing the rejection. The actual smearing is done in  $1/p_T$ , with an upper bound to the resolution uncertainty given by  $\sigma(1/p_T)=0.02/p_T$  [19]. The effect is found to be negligible.

### V. Jet transverse momentum resolution

The jet momentum resolution was measured for 2011 data and found to be in agreement with the predictions from the PYTHIA8-based simulation [40]. The precision of this measurement, determined in  $p_T$  and  $\eta$  bins, is typically 10%. The systematic uncertainty due to the calorimeter jet  $p_T$  resolution was estimated by over-smearing the jet 4-momentum in the simulated data, without changing jet  $\eta$  or  $\phi$  angles. The performance is found to globally decrease by 6%, without a particular  $p_T$  dependence.

The different contributions to the systematic uncertainty on the  $g \rightarrow b\bar{b}$  rejection are summarized in Table 6.2.

Systematic source	Uncertainty
pile-up	negligible
$b$ -tagging efficiency	negligible
track reconstruction efficiency	4%
track $p_T$ resolution	negligible
jet $p_T$ resolution	6%

Table 6.2: Systematic uncertainties in the merged  $b$ -jet rejection (common to both the 50% and the 60% efficiency working points).

## 6.5 Isolation studies

Although the tagger was derived with isolated jets it can also be applied to non-isolated jets. Studies were performed to evaluate the likelihood rejection in  $b$ -jets with close-by jet with  $p_T$  between 7 GeV at electromagnetic scale

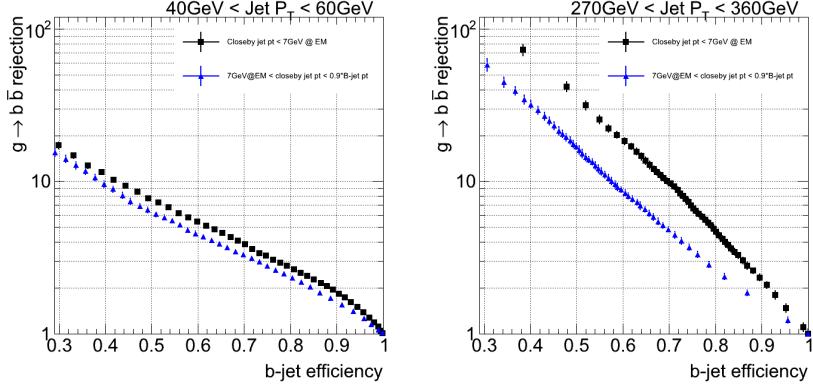


Figure 6.12: Rejection of  $g \rightarrow b\bar{b}$  merged b-jets as a function of b-jet efficiency for two different isolation cuts.

scale and 90% of the  $b$ -jet  $p_T$ . The results can be seen in Fig. 6.12. The presence of close-by jets with a substantial fraction of the  $b$ -jet  $p_T$  worsens the performance in more than 50% at very high  $p_T$ .

## 6.6 Other Monte Carlo generators

The development, training and performance determination of the tagger has been done using Monte Carlo events generated with the PYTHIA8 event simulator, interfaced to the GEANT4 based simulation of the ATLAS detector. An immediate question is what the performance would be if studied with a different simulation. In this section we investigate this question for the Perugia tune of PYTHIA8 and the HERWIG++ event generators.

Fig. 6.13 shows a comparison of the likelihood rejection, at the 50% efficiency working point, between nominal PYTHIA and the alternative simulations as a function of the jet  $p_T$ . The larger errors are due to the reduced statistics available, which are even lower for the Perugia case than for HERWIG.

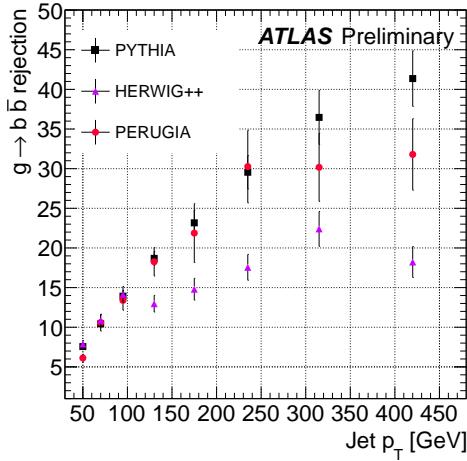


Figure 6.13: Rejection of  $g \rightarrow b\bar{b}$  merged  $b$ -jets as a function of jet  $p_T$  for different Monte Carlo generators, at the 50% efficiency working point.

The performance in HERWIG shows a systematic trend, with agreement at low  $p_T$  and increasingly poor performances compared to PYTHIA as  $p_T$  grows. For the Perugia tune, on the other hand, there is no definite behavior, with the performance fluctuating above or below the nominal simulation for different  $p_T$  bins consistently with the statistical uncertainties.

The reason for the systematic difference observed between the performances of PYTHIA and HERWIG can be traced to the extent with which jets are accurately modelled. Fig. 6.14 compares the measured jet track multiplicity distributions in  $b$ -tagged jets and the prediction from both simulations, for low and high  $p_T$  jets. It is observed that indeed HERWIG++ does not correctly reproduce the data, particularly at high  $p_T$ . The level of agreement is found to be better for track-jet width and the  $\Delta R$  between the axes of the two  $k_t$  subjets in the jet, the two other variables used for discrimination.

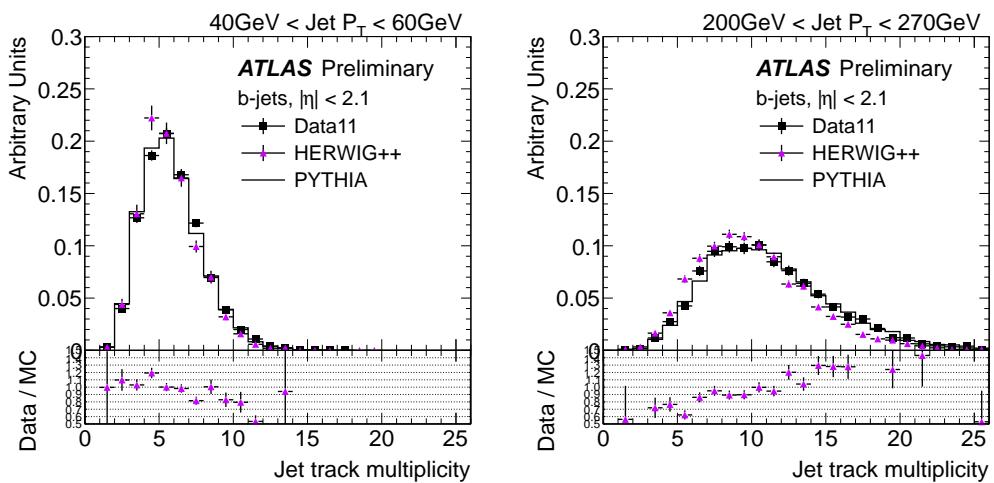


Figure 6.14: Distribution of the jet track multiplicity in 2 different jet  $p_T$  bins, for experimental data collected during 2011 (solid black points) and HERWIG++ events (solid violet triangles). The ratio data over HERWIG++ simulation is shown at the bottom of the plot. PYTHIA distribution is also shown for reference.

# Chapter 7

??Fraction of gluon-splitting  
jets in data??

## 7.1 Template fits??

# Chapter 8

## Conclusions

A multivariate discriminant to identify isolated  $b$ -tagged jets containing two  $B$ -hadrons is presented. These jets are expected to arise when a gluon splits into a close-by  $b\bar{b}$ -pair. The method exploits the kinematic differences between “merged”  $b\bar{b}$ -jets and “single”  $b$ -jets, combining track-based jet shape and jet substructure variables in a likelihood classifier.

The tagger training and performance results are based on simulated events. Several variables were investigated and those showing the best discrimination power were selected for the multivariate analysis. The Monte Carlo distributions of the explored variables were validated using experimental data corresponding to an integrated luminosity of  $4.7 \text{ fb}^{-1}$  recorded by the ATLAS experiment during 2011. The agreement between data and simulation is excellent.

The peformance of the tagger in Monte Carlo events was studied in bins of the calorimeter jet  $p_T$ , achieving rejection of merged jets of over 95% (90%) for a 50% single  $b\bar{b}$ -jet at  $150 \text{ GeV}$  ( $p_T > 60 \text{ GeV}$ ).

This tool provides a handle to investigate QCD  $b\bar{b}$  production and to reduce backgrounds in physics channels involving  $b$ -quarks in the final state.

Future improvements comprise the study of further discriminant variables, the extension to non-isolated jets using the concept of ghost-particle matching and active area of a jet [41] for track-to-jet association and labeling, the calibration of the tagger with data, and its application to measure the fraction of gluon-splitting jets in QCD  $b$ -jet production.

# Bibliography

- [1] ATLAS Collaboration. Performance of Impact Parameter-Based  $b$ -tagging Algorithms with the ATLAS Detector using Proton-Proton Collisions at  $\sqrt{s} = 7$  TeV. *ATLAS-CONF-2010-091*, 2010.
- [2] ATLAS Collaboration. Performance of the ATLAS Secondary Vertex  $b$ -tagging Algorithm in 7 TeV Collision Data. *ATLAS-CONF-2010-042*, 2010.
- [3] ATLAS Collaboration. Commissioning of the ATLAS high-performance  $b$ -tagging algorithms in the 7 TeV collision data. *ATLAS-CONF-2011-102*, 2011.
- [4] CDF Collaboration. Measurements of Bottom Anti-Bottom Azimuthal Production Correlations in Proton-Antiproton Collisions at  $s^{**}(1/2)=1.8$  TeV. *Phys.Rev.D*, 71:38, 2005.
- [5] S. Frixione and M.L. Mangano. Heavy quark jets in hadronic collisions. *Nucl.Phys.*, B483:321–338, 1997.
- [6] Andrea Banfi, Gavin Salam, and Giulia Zanderighi. Accurate qcd predictions for heavy-quark jets at the tevatron and lhc. *JHEP*, 0707:026, 2007.

- [7] S. Catani, Y.L. Dokshitzer, H. Seymour, and B.R. Webber. Longitudinally invariant  $K(t)$  clustering algorithms for hadron hadron collisions. *Nucl. Phys.*, B406:187, 1993.
- [8] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001.
- [9] John M. Campbell, R.Keith Ellis, F. Maltoni, and S. Willenbrock. Production of a  $W$  boson and two jets with one  $b^-$  quark tag. *Phys.Rev.*, D75:054015, 2007.
- [10] ATLAS Collaboration. Search for supersymmetry in pp collisions at  $\sqrt{s} = 7\text{TeV}$  in final states with missing transverse momentum,  $b$ -jets and no leptons with the ATLAS detector. *ATLAS-CONF-2011-098*, 2011.
- [11] A. Abdesselam et al. Boosted objects: a probe of beyond the standard model physics. *The European Physical Journal C - Particles and Fields*, 71:1–19, 2011.
- [12] A. Altheimer et al. Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks. 2012.
- [13] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys.Rev.Lett.*, 100:242001, 2008.
- [14] Amos Breskin and Rdiger Voss. *The CERN Large Hadron Collider: Accelerator and Experiments*. CERN, Geneva, 2009.

- [15] ATLAS Collaboration. Luminosity Determination in pp Collisions at  $\sqrt{s} = 7$  TeV using the ATLAS Detector in 2011. *ATLAS-CONF-2011-116*, Aug 2011.
- [16] S van der Meer. Calibration of the effective beam height in the ISR. *CERN-ISR-PO-68-31. ISR-PO-68-31*, 1968.
- [17] Aad, G. and others. The ATLAS Experiment at the CERN Large Hadron Collider. CERN, Geneva, 2008.
- [18] G. Aad et al. Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC. *New J.Phys.*, 13:053033, 2011.
- [19] ATLAS Collaboration. Estimating Track Momentum Resolution in Minimum Bias Events using Simulation and  $K_s$  in  $\sqrt{s} = 900$  GeV collision data. *ATLAS-CONF-2010-009*, 2010.
- [20] ATLAS Collaboration. Tracking Studies for b-tagging with 7 TeV Collision Data with the ATLAS Detector. *ATLAS-CONF-2010-070*, 2010.
- [21] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- $k_t$  jet clustering algorithm. *JHEP*, 04:063, 2008.
- [22] W Lampl et al. Calorimeter Clustering Algorithms: Description and Performance. (ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003), Apr 2008.
- [23] ATLAS Collaboration. Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data. *ATLAS-CONF-2012-020*, 2012.

- [24] ATLAS Collaboration. Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at  $\sqrt{s} = 7$  TeV and measured with the ATLAS detector. *ATLAS-CONF-2010-056*, 2010.
- [25] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. 2009.
- [26] ATLAS Collaboration. Calibrating the b-Tag Efficiency and Mistag Rate in  $35 \text{ pb}^{-1}$  of Data with the ATLAS Detector. *ATLAS-CONF-2011-089*, 2011.
- [27] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Comput. Phys. Commun.*, 178:852, 2008.
- [28] R. Corke and T. Sjöstrand. Improved parton showers at large transverse momenta. *European Physical Journal C*, 69:1, 2010.
- [29] T. Sjöstrand and P. Z. Skands. Transverse-momentum-ordered showers and interleaved multiple interactions. *European Physical Journal C*, 39:129, 2005.
- [30] B. Andersson et al. Parton fragmentation and string dynamics. *Phys. Rep.*, 97:31, 1983.
- [31] Atlas tunes of pythia 6 and pythia 8 for mc11. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011.
- [32] S. Agostinelli et al. Geant4 a simulation toolkit. *Nucl. Inst. Meth. Section A*, 506(3):250 – 303, 2003.
- [33] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur.Phys.J.C*, 70:051, 2010.

- [34] Peter Z. Skands. The Perugia Tunes. 2009.
- [35] Manuel Bahr, Stefan Gieseke, and Michael H. Seymour. Simulation of multiple partonic interactions in Herwig++. *JHEP*, 0807:076, 2008.
- [36] M Cacciari and G.P. Salam. Dispelling the  $N^3$  myth for the  $k_t$  jet-finder. *Phys. Lett. B}*, 661:057, 2006.
- [37] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with  $N$ -subjettiness. *JHEP*, 1103:015:026, 2011.
- [38] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. PoS, ACAT:040, 2007.
- [39] R Alon et al. Backup Note for Measurement of Jet Mass and Substructure in QCD with the ATLAS Experiment. ATL-COM-PHYS-2011-401, 2011.
- [40] G. Romeo, A. Schwartzman, R. Piegai, T. Carli, and R. Teuscher. Jet energy resolution from in-situ techniques with the atlas detector using proton-proton collisions at a center of mass energy  $\sqrt{s} = 7$  tev. ATL-COM-PHYS-2011-240, 2011.
- [41] G.P. Salam M. Cacciari and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 0804:42, 2008.