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

Agradezco a...

Abstract

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

Chapter 1

Introduction

The first years of proton-proton collisions at a centre of mass energy of 7 TeV delivered by the Large Hadron Collider and recorded by the ATLAS experiment have provided data to explore quantum chromodynamics (QCD) at scales never reached before. Precision measurements of strong interactions are interesting in their own right, but, in addition, QCD provides one of the main backgrounds to many New Physics measurements; furthermore, it is also through tests of QCD that New Physics may be discovered. Hadronic jets are a fundamental ingredient for precision tests of QCD: understanding and measuring their performance is crucial in the LHC environment.

A wide range of physics signatures, within the Standard Model predictions (SM) and Beyond the Standard Model (BSM), contain jets originating from bottom (b) quarks. The ability to identify jets containing b-hadrons[1] is therefore important for the high-p_T physics program of the ATLAS experiment. b-tagging algorithms rely on the relatively long decay length of

[1] Due to QCD confinement the experimental signature of quarks and gluons are not the quarks and gluons themselves but a spray of "colorless" hadrons. In the case of the b-quarks, the so called b-hadrons are observed.

b-hadrons that gives rise to large impact parameter tracks and displaced decay secondary vertices; or on the presence of a soft lepton within the jet, the product of the semileptonic b-decay. These algorithms, however, do not provide information on the number of b-hadrons within the jet. In particular, they tag jets containing a $b\bar{b}$ pair, with no net heavy flavour, which do not correspond to the intuitive picture of a b-jet as a jet containing a single b-quark or antiquark.

b-jets containing two b-hadrons, henceforth called "merged", and single b-jet, the jets containing only one b-hadron, are depicted in Fig. 1.1.

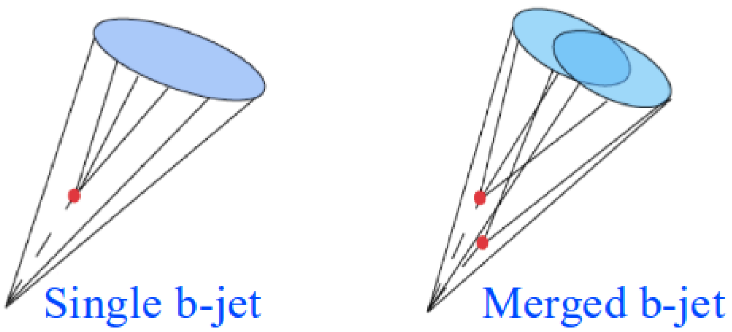

Figure 1.1: b-tagging algorithms select jets originating both from the fragmentation of single b-quark ("single" b-jet, left image) or from the splitting of a gluon into a pair of close-by $b\bar{b}$ quarks ("merged" b-jets, right image).

The ability to single out merged b-jets expected to be produced in QCD mainly by gluon splitting, has several applications. Here we briefly discuss two cases, the measurement of QCD b-quark production and the reduction of background in SM and Beyond the Standard Model (BSM) analyses.

The measurement of the inclusive b-jet spectrum

Studies of QCD bottom production are important 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, a component of the proton structure thought to be generated entirely perturbatively from the QCD evolution equations of the other flavours. The theoretical calculation of the inclusive b-jet spectrum presents rather important uncertainties ($\sim 50\%$), considerably larger than those for the light jet inclusive spectrum ($\sim 10-20\%$) [1]. These arise from the poor convergence of the perturbative series, as evidenced by the large value of the K-factor (NLO/LO), $K = 6 - 10$, in the p_T range covered by the LHC. While at LO only the so-called "flavour creation" channel is present, at NLO two new channels open up, often referred to as "flavour excitation" (FEX) and "gluon splitting" (GSP), see Fig. 2.1. NLO effects are included approximately in LO parton showering models, such as Herwig or Pythia. The various channels can be approximately separated in a parton shower Monte Carlo generator such as these, where one can determine the underlying hard process from the event record. It is found that the LO channel has a much smaller contribution than the FEX and the GSP channels, which receive strong enhancement from collinear logarithms [?].

Ref. [2] proposes a new observable to free the heavy-flavour spectrum calculation from collinear logarithms, and improve the accuracy of the theoretical prediction, by not including in the production cross-section the contribution from double b-hadron jets. Final-state logarithms are removed by employing a recently developed jet reconstruction scheme, the flavour-k_t algorithm [3], which maintains the correspondence between partonic flavour and jet flavour. Specifically, jets containing a b-quark and a b-antiquark, which in a parton shower MC generator are produced $\sim 95\%$ of the time by the GSP channel, are labeled in an IR-safe way as light jets and removed from the b-jet spectrum. The initial-state (FEX) collinear logarithms can be resummed by using a b-quark parton distribution functions. With this algorithm the K-factor for the differential heavy-jet spectrum cross-section is shown not to exceed a value of $K = 1.4$, with a factor of four reduction in the theoretical (scale variation) uncertainties.

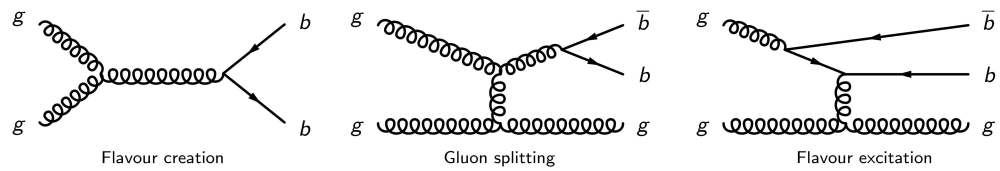

Figure 1.2: Representative diagrams of the three channels contributing to QCD b-quark production up to NLO. The flavour creation channel (left) is the only one present at LO. At NLO, two new channels open up, referred to as gluon splitting (center) and flavour excitation (right).

Rejection of background in Standard Model analyses and beyond-SM searches

Efficient tagging of merged b-jets can provide an important handle to understand, estimate and/or reject b-tagged backgrounds to SM and new physics searches at the LHC.

SM physics analyses that rely on the presence of single b-jets in the final state, such as top quark physics (either in the $t\bar{t}$ or the single top channels) or associated Higgs production ($WH \to \ell\nu b\bar{b}$ and $ZH \to \nu\nu b\bar{b}$), suffer from backgrounds that can be in part removed by a merged b-jet tagger. These are the reducible background from QCD, which can produce double b-hadron jets as discussed in the previous subsection, and the irreducible background due to W bosons plus b-jets. While at LO only single b-jets are present in $W + b$ production, at NLO jets containing two b-hadrons are expected due to the contribution of diagrams containing a $gb\bar{b}$ vertex. The relevance of merged b-jets in this channel is supported by NLO calculations of the production of W bosons and two jets with at least one b-quark at the LHC [4]. For jets with $p_T > 25$ GeV, and $|\eta| < 2.5$, they indicate that the cross section for $W(b\bar{b})j$ is almost a factor of two higher than $Wb\bar{b}$, where $(b\bar{b})$ denotes the case in which the two b-quarks are merged into the same jet.

Jets containing a single b-quark or antiquark also enter in many BSM collider searches, notably because b-quarks are produced in the decays both of heavy SM particles, top quarks, the Z boson and the Higgs boson, and of particles appearing in proposed extensions of the SM. The ability to distinguish single b-jets from jets containing two b-hadrons is thus here of wide application to reduce SM backgrounds giving rise to close-by $b\bar{b}$ pairs.

There are two possible strategies to attempt to identify b-jets containing two b-hadrons. One of them relies on the direct reconstruction of the two b-decay secondary vertices [5]. This allows the measurement of the angular separation between the b-hadrons, but suffers from the low efficiency of a double b-tag requirement plus additional reconstruction inefficiencies at small angular separation between the two b-hadrons. In this thesis we develop an alternative method that does not rely on explicit vertex finding, but exploits the substructure differences between single and merged b-jets, combining them in a multivariate analysis.

Chapter 2 describes the theoretical framework, with emphasis in the theory of the strong interactions and the aspects that are important for the understanding of the hadronic final state in hadronic collisions. The LHC and the ATLAS detector components are described in Chapter 3, together with a summarization of the detector conditions during 2011 data taking. Chapter 4 details how jet reconstruction and calibration are performed at ATLAS and describes the procedure for the identification of b-quark jets. Chapter 5 presents the analysis of jet shape and substructure variables for the discrimination between single and double b-hadron jets. The validation of the variables in 2011 data is also included. The construction of the multivariate discriminator and the discussion of the systematic uncertainties are presented in Chapter 6.

TO DO Preliminary results for the measurement of the fraction of double b-hadron jets in data are discussed in Chapter 7.

Finally, chapter 8 summarizes the results.

Chapter 2

Theoretical framework

In this chapter a short overview of the theory of elementary particles and fundamental interactions is presented, with emphasis on the strong interactions and the description of the hadronic final state in hadron collisions.

2.1 The Standard Model

The Standard Model (SM) is a quantum field theory that describes the behavior of all experimentally-observed particles under the influence of the electromagnetic, weak and strong forces[1]. In this model, all forces of nature are the result of particle exchange. The force mediators interact on the particles of matter, and, in some cases, due to the non-Abelian character of the theory[2], with each other.

The fundamental building blocks of matter predicted by the theory are

[1]In principle gravitational forces should also be included in the list of fundamental interactions but their impact is fortunately negligible at the distance and energy scales usually considered in particle physics experiments.

[2]The transformations of the symmetry group do not commute in the case of the QCD and weak groups.

fermions with spin 1/2:

- six leptons (and their antiparticles), organized in three families,

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}$$

- and six quarks (and their antiparticles), organized in three families,

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$$

.

The six types of quark are also known as the six quark flavors. Collectively, the u (up), d (down), and s (strange) quarks are frequently referred to as the light quarks. The heaviest quark of the Standard Model, the quark t (top), was the last to be found [6] [7]. These particles are considered point-like, as there is no evidence of any internal structure of leptons or quarks to date.

In addition, the model contains the vector bosons which are the carriers of the the fundamental forces:

- a gauge boson for the electromagnetic interactions, the photon γ;
- three gauges bosons for the weak interactions, $W^\pm$ and Z^0;
- eight gauge bosons for the strong interactions, called gluons.

The exact symmetry (see below) of the SM predicts massless particles. One possible mechanism for breaking this symmetry is the existence of a massive scalar Higgs field that has non-zero vacuum expectation value [8]. Very recently, a Higgs-like particle was discovered by ATLAS and CMS experiments at the LHC [9, 10]. This scalar boson completes the table of Standard Model particles.

Gauge invariance, defined as the invariance of the theory under local transformations, is a fundamental property of the SM. In the theory, the electromagnetism (Quantum electrodynamics), the weak interaction, and the strong nuclear force (Quantum Chromodynamics) are all derived from imposing Lorentz invariant symmetries onto the interacting fields.

The theory of Quantum electrodynamics (QED) describes the interaction of charged particles via the exchange of one (or more) photon. It is formulated by imposing a $U(1)$ or rotational symmetry onto the simplest field lagrangian that obeys the correct equation of motion. The full theory of QED was developed by Richard Feynman and others throughout the 1940s [11]. The structure of the SM is in a sense a generalisation of this theory, extending the gauge invariance of electrodynamics to a larger set of conserved currents and charges.

The symmetry associated to the weak interaction is the $SU(2)$ symmetry, which corresponds to rotations of 2-dimensional vectors. The latter combines with the $U(1)$ symmetry from QED to produce additional gauge fields. The gauge fields merge with the gauge field from QED to form W^+, W^- and Z^0 bosons that are the carriers of the weak force. Unlike the photon, which is massless, the $W^\pm$ and Z^0 bosons have masses close to 80 and 90 GeV, respectively [12] [13]. Due to these large masses, the weak force has a short range and is feeble at low energies. At masses higher than the Z mass, the electromagnetic and weak forces unify into a single force, known as the electroweak force [14] [15] [16].

The current theoretical theory of the strong interactions began with the identification of the elementary fermions that make up the proton and other hadrons. In 1963, Gell-Mann and Zweig propose the quark model [17] [**?**] [18], which asserts that these particles are in fact composites of smaller constituents. Mesons were expected to be quark-antiquark bound states and Baryons were interpreted as bound states of three quarks, all with fractional charges. The quark model was formalized into the theory of Quantum Chromodynamics (QCD) by Harold Fritzsch and Murray Gell-Mann [?] [19] in 1973, who proposed that quarks carried an additional quantum number called color. Whitout color charge, it would seem that the quarks inside some hadrons exist in identical quantum states, in violation of the Pauli exclusion principle (this was indeed the problem of the quark model as proposed by Gell-Mann and Zweig). The color theory extends the electroweak Lagrangian to be symmetric under $SU(3)$ transformations, which introduces eight new physical gauge fields, the gluons. Due to Richard Feynman's parton model nomenclature [20], both quarks and gluons are commonly referred to as partons.

Another problem of the quark model was that free particles with fractional charges were never found. The answer to why we never see free quarks or gluons outside of a hadron, together with the tools for performing theoretical calculations in QCD are given the next sections.

2.2 Perturbative QCD

As described above, the fundamental actors of the theory of the strong interactions are quarks and gluons or, collectively, partons. Partons are confined in hadron like particles, like the proton, but, act free at sufficiently small scales. This behaviour is called asymptotic freedom. The essence of asymptotic freedom is that the strong force couples particles together more strongly as the distance between them increases. The experimental consequence of asymptotic freedom is that quarks and gluons require interactions with high energy probes to be ejected from nucleons, and they cannot be observed directly.

First indications of the presence of quarks resulted from the measurement of deep inelastic lepton-hadron scattering. The momentum transfer, Q^2, between the probe particles (leptons) and the target hadron is analogous to the distance scale within the hadron being measured. The variation of the strength of the coupling with the energy is referred to as the "running" of coupling constant.

The low value of the strong coupling constant at high-energies permits the use of perturbative techniques to calculate physical processes. As one goes higher in the perturbative expansion, each term contains an additional factor of the coupling constant, α_s. Since α_s depends on the energy, it must be evaluated at some energy scale, close to the energy scale involved in the process. For instance, at an energy of 15 GeV, the strong coupling constant takes on an approximate value of 0.1, thus, from an expansion of an infinite number of terms, only a few need to be computed. The complexity of the process determines the precision of the calculation that can be permormed. For example, predictions for the cross section for events with three partons in the final state are only available up to leading-order (LO). For inclusive parton production, calculations are typically performed at next-to-leading order (NLO). Feynman diagrams are used in the computation of the multiple terms in the expansion, they are graphical representations of each term. Example of Feynman diagrams for QCD b-quark production up to NLO are shown intance Fig. 2.1.

With this formalism, the cross-section for the interaction of partons can be computed up to fixed-order in perturbation theory. In the case of hadron colliders, such as the LHC, is the factorization theorem [21] that allows the

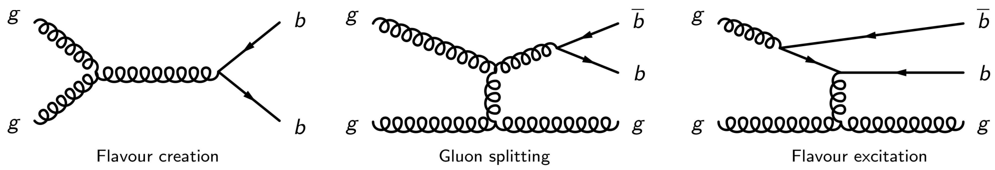

Figure 2.1: Representative diagrams of the three channels contributing to QCD b-quark production up to NLO.

perturbative calculations for parton interactions to be extended to proton-proton collisions.

In the simplest picture, a proton is a combination of three quarks: two up quarks and one down quark. The reality, however, is much more complex. In the proton there are gluons constantly being emitted and absorbed, causing quark/antiquark pairs of many flavors to be briefly produced and destroyed. The up and down quarks of the standard hadron model are called valence quarks, while the virtual quark/antiquark pairs are known as sea quarks. Both valence quarks and sea quarks, along with the gluons, share the total momentum of the hadron. The distribution of the momentum fraction, x, carried by each parton is expressed as a probability to find a particular parton with a given x. The latter is known as the Parton Distribution Function (PDF).

For cross-section calculations, the PDFs are evaluated at a factorization scale, μ_f, which can be thought of as the scale that separates short-distance, perturbative physics from long-distance, non perturbative physics. Any variation in the computed cross sections due to different choises of the energy scales can be interpreted as an uncertainty due to the unknown higher-order corrections in the cross-section calculation.

The evolution of the PDF on Q is given by the DGLAP equations, published separately in the 1970s by Yuri Dokshitzer, Vladimir Gribov and Lev

Lipatov, and Guido Altarelli and Giorgo Parisi [22]. The DGLAP equations are derived by noting that the PDFs should be independent of the factorization scale f . This gives a precise mathematical form to the dependence. The dependence on x, however, must be obtained by fitting possible cross section predictions to data from hard scattering experiments.

2.3 Jet physics

Due to confinement the experimental signature of quarks and gluons are the final state "coloreless" hadrons[3]. The packet of particles produced tends to travel collinearly with the direction of the initiator quark or gluon. The result is a collimated "spray" of hadrons (also photons and leptons) entering the detector in place of the original parton; these clusters of objects are what we define as jets. The first evidence for jet production was observed in e^+e^- collisions at the SPEAR storage ring at SLAC in 1975 [23].

The evolution from a single parton to an ensemble of hadrons accurs through the processes of parton showering and hadronization. Since the strong coupling constant grows with increasing distance between color charges, a strong color potential forms as the parton from the "hard" (high Q^2) scattering process separates from the original hadron. This large potential causes quark/antiquark pairs ($q\bar{q}$) to be created, each carrying some of the energy and momentum of the original partons. As these new partons move away from one another, yet more color potentials are formed, and the process repeats. Thus from one parton a shower of partons appears, traveling along the same direction as the original. This process continues until there is no longer enough energy to create additional $q\bar{q}$ pairs, and instead the remaining

[3]We use "colorless" to mean a singlet representation of the color group.

partons combine to form stable hadrons. Since this progression involves successively lower energies and lower momentum transfers, perturbative QCD cannot describe the full process. The hadronization process then cannot be calculated from first principles, but has to be modelled.

2.3.1 Monte Carlo tools

Knowing QCD predictinos is crucial in the design of methods to search for new physics, as well as for extracting meaning from data. Different techniques can be used to make QCD predictions at hadron colliders, and in particular at the LHC. The so called Matrix Element Monte Carlos use direct perturbative calculations of the cross-section matrix elements for each relevant partonic subprocesses. LO and NLO calculations are available for many processes. These "fixed-order predictions" include the first terms in the QCD perturbative expansion for a given cross-section; as more terms are involved in the expanstion, an improvement in the accuracy of the prediction is expected. The complexity of the calculations increased significantly with the number of outgoing legs, limiting available results to those with at most three outgoing partons. Matrix element MC programs include Alpgen [24], Madgraph [25] and others.

An alternative approach is applied by the so called Monte Carlo parton shower programs. These simulation programs use LO perturbative calculations of matrix elements for $2 \to 2$ processes, relying on the parton shower to produce the equivalent of multi-parton final state. Pythia [26] and Herwig++ [27] are the most commonly used parton shower Monte Carlos together with SHERPA.

The Monte Carlo generators must account for and correctly model the showering of partons. To approximate the energy-evolution of the shower, the DGLP equations that describe the evolution of the PDFs with changing energy scale can be used. The separation of radiation into initial- (before the hard scattering process takes place) and final-state showers is arbitrary, but sometimes convenient. In both initial- and final-state showers, the structure is given in terms of branchings $a \to bc$: $q \to qg$, $q \to q\gamma$, $g \to gg$ and $g \to q\bar{q}$. Parton b carries a fraction z of the energy of the mother energy and parton c carries the remaining $1 - z$ (the term "partons" is including the radiated photons). In turn, daughters b and c may also branch, and so on. Each parton is characterized by some evolution scale, which gives an approximate sense of time ordering to the cascade. In the initial-state shower, the evolution scale values are gradually increasing as the hard scattering is approached, while these values decrease in the final-state showers. The evolution variable of the cascade in the case of PYTHIA generator, Q^2, has traditionally been associated with the m^2 of the branching partons[4]. In the recent version of PYTHIA a $p_\perp$-ordered shower algorithm, with $Q^2 = p_\perp^2$ is available, and the shower evolution is cut off at some lower scale Q_0 typically arond 1 GeV for QCD branchings. HERWIG++ provides a shower model which is angular-ordered.

There are two leading models for the description of the non-perturbative process of hadronization, after parton showering. PYTHIA uses the Lund string model of hadronization to form particles [28]. This model involves stretching a colour "string" across quarks and gluons and breaking it up into hadrons.HERWIG++ utilizes the cluster model of hadronization. In this model each gluon is split into a $q\bar{q}$ pair and then quarks and anti-quarks are grouped into colourless "clusters", which then give the hadrons.

[4]The final-state partons have $m^2 > 0$. For initial-state showers the evolution variable is $Q^2 = -m^2$, which is required to be strictly increasing along the shower.

Hadronization models involve a number of non-perturbative parameters. The parton-shower itself involves the non-perturbative cut-off Q_0^2 . These different parameters are usually tuned to data from the LEP experiments.

In addition to the hard interaction that is generated by the Monte Carlo simulation, it is also necessary to account for the interactions between the incoming proton remnants. This is usually modelled through multiple extra $2 \to 2$ scattering, occurring at a scale of a few GeV. This effect is known as multiple parton interactions (MPIs). In addition, these partons may radiate some of their energy, either before or after the hard interaction. All the parton interactions, which are note calculated form the hard scattering process, are grouped together in the term underlying event. The modelling of the underlying event is crucial in order to give an accurate reproduction of the (quite noisy) energy flow that accompanies hard scatterings in hadron-collider events.

It should be stressed that these multiple parton interactions are a separate effect from the multiple proton interactions that may occur in each collision event in the LHC. These multiple proton collisions are referred to as pileup, and are not included in the definition of the underlying event.

No precise model exists to reproduce the unerlying event activity. This activity is instead also adjusted to reproduce available experimental data. A specific set of chosen parameters for a generator is referred to as a "tune".

The two Monte Carlo generators used in this analysis are summarized below, indicating the particular versions and tunes that were implemented.

Pythia

Pythia event generator has been used extensively for e^+e^-, ep, $pp/p\bar{p}$ at LEP, HERA, and Tevatron, and during the last 20 years has probably been the most used generators for LHC physics studies. PYTHIA contains an extensive list of hardcoded subprocesses, over 200, that can be switched on individually. These are mainly 2→1 and 2→2, some 2→3, but no multiplicities higher than that. Consecutive resonance decays may of course lead to more final-state particles, as will parton showers.

As mentioned above, in this MC generator, showers are ordered in transverse momentum [29] both for ISR and for FSR. Also MPIs are ordered in p_T [30]. Hadronization is based solely on the Lund string fragmentation framework.

For the results presented in this thesis simulated samples of dijet events from proton-proton collision processes were generated with P YTHIA 6.423 [26]. The ATLAS AMBT2 tune of the soft model parameters was used [31]. This tune attemps to reproduce the ATLAS minimum bias charged particle multiplicity and angular distribution measurements and the ATLAS measurements of charge particle and p_T density obseved collinear and transverse to the high-energy activity.

For systematic comparisons, a set of additional tunes, called the Perugia tunes [32] were also used. These tunes utilize the minimum bias and p_T density measurements of CDF to model the underlying event, hadronic Z^0 decays from LEP to model the hadronization and final state radiation, and Drell Yann measurements from CDF and $D0$ to model the initial state radiation. In particular, the Perugia 2011, which is a retune of Perugia 2010 [33] including 7 TeV data (Mar 2011).

Herwig++

HERWIG++ [27] is based on the event generator HERWIG (Hadron Emission Reactions With Interfering Gluons), which was first published in 1986 and was developed throughout the era of LEP. HERWIG was written in Fortran, and the new generator, Herwig++ developped in C++. Some distinctive features of Herwig++ are: Angular ordered parton showers and cluster hadronization. Hard and soft multiple partonic interactions to model the underlying event and soft inclusive interactions [?].

This MC generator was used for systematic uncertainties studies. The version utilized was versin 2.4.2 released in 2009.

In order to use events produced by Monte Carlo generators to model events that one might observe with the detector, the output of these generators is passed through a detector simulation model. ATLAS uses the GEANT4 [34] toolkit to simulate the passage of particles through the detector material. This includes models for the production of additional particles caused by inelastic scattering off of electrons and nuclei, as well as ionization and absorption by active detector elements.

2.3.2 Jet algorithms

As described above, quarks and gluons cannot be directly observed. Almost inmediatly after being produced, a quark or gluon hadronises, leading to a collimated spray of energetic hadrons, a jet. By measuring the jet energy and direction one can close to the idea of the original parton. But one parton may form multiple experimentally observed jets, for example due to a hard gluon emission plus soft and collinear showering. Then, in comparing data to theory and MC programs predictions a set of rules for how to group particles into jets is needed. A jet algorithm, together with a set of parameters and a recombination scheme (how to assign a momentum to the combination of two particles) form a jet definition.

By using a jet definition a computer can take a list of particle momenta for an event (be they quarks and gluons, or hadrons, or calorimeter depositions), and return a list of jets. One important point to remark is that the result of applying a jet definition should be insensitive to the most common effects of showering and hadronization, namely soft and collinear emissions. This is illustrated in Fig. 2.2.

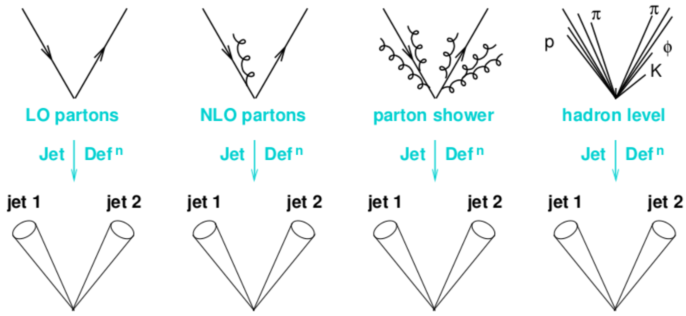

Figure 2.2: The application of a jet definition to a variety of events that differ just through soft/collinear branching and hadronization should give identical jets in all cases [35].

Tradiontally, jet algorithms have been classified into two categories: cone algorithms and sequential recombination algorithms.

Cone-like algorithms are based on the collinear nature of gluon radiation and the parton shower described above. The decay products of and emission from a hard quark or gluon will tend to form a cone of particles in the $\eta - \phi$ plane as they propagate. An cone algorithm will work as follows [5]: first, it sorts all particles in the event according to their momentum, and identifies the one with largest p_T. This is referred to as seed particle. Then a cone of radius R in $\eta - \phi$ is drawn around the seed. The direction of the sum of the momenta of those particles is identified and if it doesn't coincide with

[5]This is how CMS cone algorithm, used for the preparation for the LHC running, works.

the seed direction then the sum is used as a new seed direction, and iterates until the sum of the cone contents coincides withthe previus seed (this type of algorithm is cone "iterative" cone since it iterates the cone direction). This is how a stable cone is reached. A difficulty and major drawback in this procedure is the use of the transverse momentum of the particle to select the first seed. This definition is collinear unsafe, i.e. a splitting of the hardest particle into a nearly collinear pair can have the consequence that another, less hard particle, pointing in a different direction suddenly becomes the hardest in the event, leading to a different final set of jets. There are many other variants of cone algorithms, and nearly all suffer from problems of either collinear safety, or infrared safety (an extra soft particle creates a new seed, which can lead to an extra stable cone being found). A fix for these problems came in a algorithm called Seedless Infrared Safe Cone (SISCOne) [36].

Recombination algorithms, on the other hand, are both collinear and infrared safe. And for this reason, they can be used in calculations to any order in perturbation theory. The term recombination is used given that these algorithms work as if they were inverting the sequence of splittings of the parton shower. In general, recombination algorithms operate by successively combining pairs of particles using a distance metric, d_{ij}. At hadron colliders, due to the fact that one of the incoming partons may continue along the beam, for every pair of particles this metric is compared to a so-called "beam distance", d_{iB}, and only when $d_{ij} < d_{iB}$ the particle pair is combined and considered for subsequent clustering steps.

ATLAS (and also CMS) has chosen anti-k_t [37] algorithm as the default jet algorithm for use in physics analysis. This recombination algorithm as well as the Cambridge-Achen algorithm [?], or C/A are extensions of the original k_t algorithm developed for the analysis of multi-jet events at e^+e^-

colliders [38] and subsequently extended for use at hadron colliders [39] [40]. In this thesis, the k_t algorithm was used for jet substructure studies, see section 2.3.3.

The orginal k_t algorithm implements the following (2.1) distance metric between particles i and j,

$$d_{ij} = \frac{2E_iE_j(1 - cos\theta_{ij})}{Q^2} \tag{2.1}$$

where Q is the total energy in the event, E_i is the energy of particle i and θ_{ij} the angle between particles i and j. In the collinear limit, d_{ij} is related to the relative transverse momentum between particles i and j (hence the name k_t algorithm), normalized to the total visible energy. The particles are combined if the minimum d_{ij}, d_min, is below a certain threshold, y_{cut}. The jet multiplicity depends on the value of y_{cut}, as a lower value will result in more soft or colliinear emissions surviving as jets. As mentioned above, for hadron colliders, the notion of a beam distance is added. A distance scale, $\Delta R = \sqrt{\Delta y^2 + \Delta\phi^2}$, is introduced to define the typical radius for a jet, effectively replacing y_{cut}. In this case for every pair of particles a new distance is define, (2.2),

$$d_{ij} = min(p_{ti}^2, p_{tj}^2)\frac{\Delta R_{ij}^2}{R^2} \tag{2.2}$$

and the beam distance, $d_{iB} = p_{ti}^2$. The algorithm proceeds by searaching for the smallest of the d_{ij} and the d_{iB}. If it is a d_{ij} then particles i and j are recombined into a single new particles. If it is a d_{iB} then i is removed from the list of particles, and called a jet. This is repeated until no particles remain.

As opposed to cone algorithms, for the k_t algorithm, the jets have quite irregular shapes, and particles with $\Delta R_{ij} > R$ can still be clustered within the jet. This is a problem when, for example, an irregularly shaped jet happens to extend into poorly instrumented detector regions. Another drawback of this definition is that soft particles are clustered first. This has the potential to introduce complications when the detector noise of energy density fluctuations are large.

A feature of the k_t algorithm that is attractive is that it not produces jets but also assigns a clustering sequence to the particles within the jet. It is possible then to undo the clustering and look inside the structure of the jet. This has been exploited in a range of QCD studies, and also in searches of hadronic decays of boosted massive particles and will be used here for the search of two-pronged jets in gluon splitting.

The prescription above may be generalized beyond the k_t algorithm. By inverting the power law in the particle distance metric, d_{ij}, the anti-k_t algorithm is obtained. The particle distance metric used by this algorithm is,

$$d_{ij} = min(p_{ti}^{-2}, p_{tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \tag{2.3}$$

and the beam distance, $d_{iB} = p_{ti}^{-2}$. This definition results in the clustering of the hardest emissions first. This has several benefits in the context of high-luminosity hadron collisions.

Note that the anti-k_t algorithm does not provide useful information on jet substructure if a jet contains two hard cores, then the k_t (or C/A) algorithms first reconstruct those hard cores and merge the resulting two subjets. The anti-k_t will often first cluster the harder of the two cores and then gradually aglomerate the contents of the second hard core.

These algorithms, and more, are implemented in FASTJET [41] software package for jet-finding.

2.3.3 Jet substructure

The study of a quantity related to the distributions and multiplicity of particles in the event phase space lead to the first evidence of jet structure, as pointed out in ref. [23]. In general, all final hadronic states in $pp/p\bar{p}/e^+e^-$ collisions can be explored in terms of the structure and shape of the event energy flow by means of so called "event shape" variables. This family of variables attempt to extract information about the global geometry of an event, usually distinguishing between di-jet events and multijet final states. Such variables have been successfully utilized in many SM measurements and BSM searches, see for example [42][43].

Although very useful, event shape variables are not sensitive to the detailed structure and distribution of energy inside a particular jet in the event. In new physics searches, tools for the identification of individual objects that might be signature of new particles are desired. For example, when an unstable massive particle with large p_T decays hadronically, the final state may be composed of a number of nearly collinear jets. These jets may be merged by a jet finder; a method for selecting these jets would allow for the study of their properties. This interest lead to the development of a wide range of jet substructure techniques in the recent years.

Jet substructure methods probe the internal structure of jets from a detailed study of its constituents (see chapter**??**). These techniques have been first thought for distinguishing boosted hadronic objects from the background of jets initiated by light quarks and gluon, see for example [44], but they have been also used succesfully in other applications, including separating quark jets from gluon jets [45] and identifying boosted decay producs in new physics [46].

Jet shapes, which are event shape-like observables applied to single jets, are an effective tool to measure the structure of individual jets [47].The shape of a jet no only depends on the type of parton (quark or gluon) but is also sensitive to non-perturbative fragmentation effects and underlying event contributions [48].

In the particular case of the present analysis, several distinguishing characteristics between jets originating from b-quarks and jets originating from the the splitting of a gluon into a $b\bar{b}$ pair can be determined using the techniques of jet substructure.

Jet width

The jet width is part of a set of continuous variables that try to distinguish individual particles/subjets within the jet as a smooth funcion of $(\delta\eta, \delta\phi)$ away from the jet axis, in order to form combinations like geometric moments. This particular combination sums the distances between the jet constituents and its axes, weighted by the constituent p_T, and then normalized to the total p_T of the jet. The compact definition is

$$Jet\ width = \frac{\sum_{i=1}^{N} {p_T}^{const_i}\, \Delta R(const_i, jet)}{\sum_{i=1}^{N} {p_T}^{const_i}} \tag{2.4}$$

where N is the total number of calorimeter or track constituents. This observable is also highly correlated to the mass of the jet.

This linear radial moment is a measure of the width or “girth” [49] of the jet. Under the assumption of central jets with massless constituents at small angles, this linear moment is identical to jet broadening [50], defined as the sum of momenta transverse to the jet axis normalized by the sum of momenta. While jet broadening is natural at an e^+e^- collider, the linear radial moment is more natural at the LHC.

An alternative approach to measuring the width is to use the angular separation of the two hardest constituents inside jets. This has the advantage of effectively removing any dependence on the shower development within the calorimeter and focuses on the hard component of the jet.

Eccentricity

In defining a jet moment there are several ways to weight the momentum and define the center of the jet. We have defined the jet width as the first moment of the transverse energy with respect to the jet axis; another example of useful combination is the jet pull [49]. But it is also natural to look at higher moments, such as those contained in the covariance tensor,

$$C = \sum_{i \in jet} \frac{p_T^i |r_i|}{p_T^{jet}} \left(\begin{array}{cc} \Delta y_i^2 & \Delta y_i \Delta\phi_i \\ \Delta\phi_{ii} & \Delta\phi_i^2 \end{array} \right).$$

Here, $r_i = (\Delta y_i, \Delta\phi_i) = c_i - J$, where $J = (y_J, \phi_J)$ is the location of the jet and c_i is the position of a cell or particle with transverse momentum p_T^i. The eigenvalues $a \geq b$ of this tensor are similar to the semimajor and semiminor axes of an elliptical jet. The jet eccentricity, defined below, is a combination of these eigenvalues, and it is a measure of how elongated is the area of a jet.

$$e = \sqrt{\frac{(a^2 - b^2)}{a}} \tag{2.5}$$

Jet Mass

The jet mass, like the linear radial moment, also depends on the radiation pattern of the event. It is the most basic observable for disinguishing massive boosted objects from jets originating from quarks or gluons. The latter are expected to be dominated by wide-angle emissions, with increase probability to see high mass jets initiated from gluons as opposed to quarks [51].

Need to complete this.

Subjet multiplicity

With the development of the k_t algorithm, subjets were first used in the description of the hadronic final state in e^+e^- annihilation, such as the study of the jet multiplicity at different energy scales [52]. By using the sequential recombination algorithms introduced in the previus section, it is straightforward to define a “subjet algorithm” in which the structure of the jet’s constituents is resolved using either the same jet finder algorithm or a new one with a fixed (smaller) distance parameter.

The subjet multiplicity − the number of subjets within a jet − provides information on the distribution of energy and multiplicity of particles within a jet. For instance, in [53] the result of meassuring this “radiation variable” on quark- and gluon-initiated jets indicates that gluon-initiated jets tend to have on average higher subjet multiplicity. This result is consistent with the QCD prediction that gluons radiate more than quarks. In the case of this and different other analyses the k_t algorithm is rerun for subjet finding.

As an alternative to fixed distance parameter subjets, it is also possible to undo the last step in the recombination sequence [39] in order to identify the decay products of an object. This approach is used in seveal jet grooming procedures[6], see for instance [55].

It is also possible to extend the use of individual subjets in conjunction with more traditional jet shape variables. Using these tools, an inclusive jet shape based on the substructure topology of a single jet, “N-subjettiness” [56] is defined.

[6] Jet grooming comprises dedicated techniques to remove uncorrelated radiation within a jet. A review of these procedures can be found in [54].

N-subjettiness

As mentioned above, the N-subjettiness [56] is a jet shape that describes the energy flow within a jet. It quantifies the degree to which radiation is aligned along specified subjet axes. This jet shape was adapted from the event shape N-jettiness [57].

Given candidate subjets directions determined by an external algorithm such as the exclusive k_t procedure [58], the variables is defined as,

$$\tau_N^{(\beta)} = \frac{1}{\sum_k p_{Tk}\,(R_0)^\beta} \sum_k p_{Tk} (\min\{\Delta R_{j1,k},\ \Delta R_{j2,k}, ...,\ \Delta R_{jN,k}\})^\beta \qquad (2.6)$$

The sum runs over the k constituent particles in a given jet where $p_{T,k}$ are their transverse momenta, and $\Delta R_{j1,k}$ is the distance between the candidate subjet $j1$ and a constituent particle k. R_0 is the characteristic jet radius used in the original jet clustering algorithm. The exponential weight, β, can optionally be applied to the angular distance computed between the subjets and the jet constituents.

This jet shape was designed to identify boosted N-prong hadronic decays. With $\beta = 1$, the definition above indicates that jets with $\tau_N \approx 0$ have all their radiation aligned with the candidate subjet directions and therefore have N (or fewer) subjets. Jets with $\tau_N \gg 0$ have a large fraction of their energy distributed away from the candidate subjet direction and therefore have at least $N + 1$ subjets.

To separate boosted hadronic objects from the QCD jet background, one could use the complete set of τ_N (with different values of β) in a multivariate analysis. However, [56] showed that a simple cut on the ratio τ_N/τ_{N-1} provides excellent discrimination power for N-prong hadronic objects. In particular, τ_2/τ_1 can identify boosted W/Z and Higgs bosons, with the angular weighting exponent $\beta = 1$ providing the best discrimination.

Since eq. 2.6 is linear in each of the constituent particle momenta, this variable is an infrared- and colliner-safe observable. In subsequent work [59], Thaler and van Tilburg showed that the initial step of choosing candidate subjet axes is in fact unnecessary. In particular, the quantity in equation 2.6 can be minimised over the candidate subjet directions, further improving boosted object discrimination.

The definition of N-subjettiness is not unique, and different choises can be sued to give different weights to the emissions within a jet. There generalizations of N-subjettiness are similar to different "angularities" [60] used in $e^+e^- \rightarrow$hadrons measurements.

Chapter 3

The ATLAS detector at the LHC

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [61] 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.

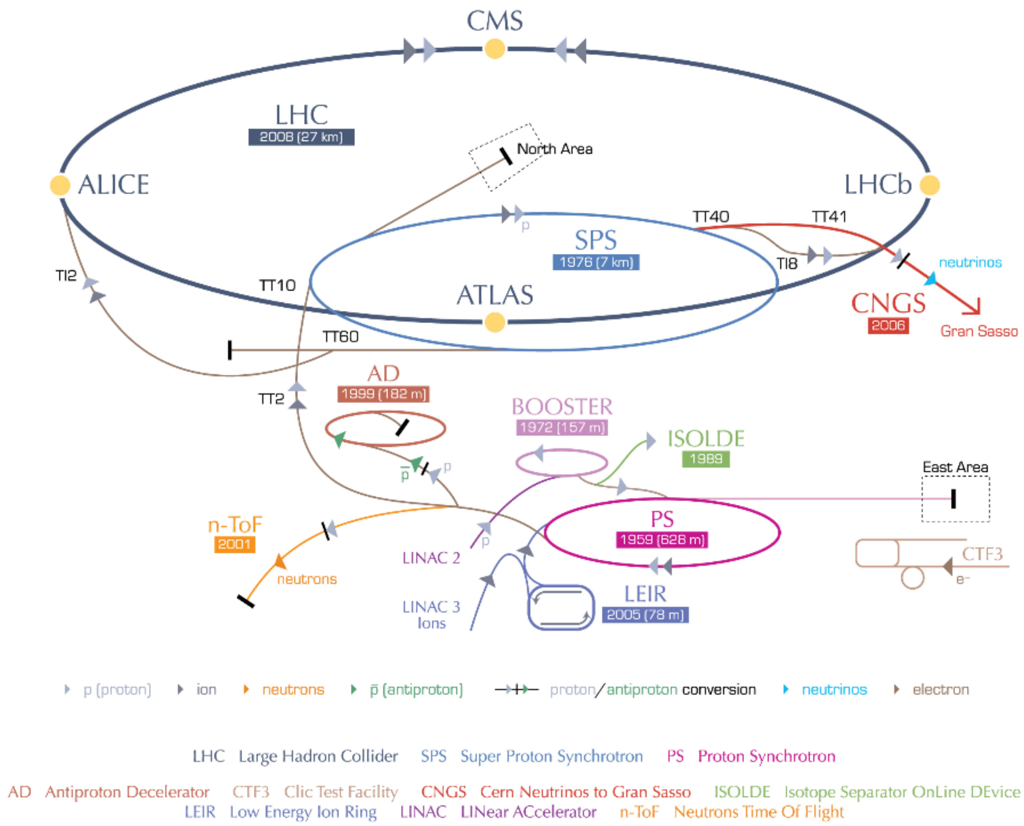

Figure 3.1: The CERN accelerator complex, showing the injection system, along with each components date of construction, and the placement of the four main experiments.

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 Synchroton (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 intensitity, 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 ocurred on November 23 2009, at the SPS extraction energy of 450 GeV per beam. Very quick 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 1.5×10^{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 ≈ 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
Instantaneus luminosity [$\mathrm{cm}^{-2}\mathrm{s}^{-1}$]	$3.65\ 10^{33}$ (year peak)	10^{34}
Bunches per beam	38 (May)	2808
Protons per bunch	0.8×10^{11} (May)	1.5×10^{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 $\mathrm{cm}^2\mathrm{s}^{-1}$. The number of events of a particular process, then, is given by the product of the integrated luminosity, $\int dtL$, and the cross section of the process, σ_{event}. The integrated luminosities are typically quoted in units of inverse picobarns, $\mathrm{pb}^{-1} = 10^{-36}\mathrm{cm}^2$. In order to measure processes with very little cross sections a very high luminosity is required.

The delivered luminosity can be written as [62]:

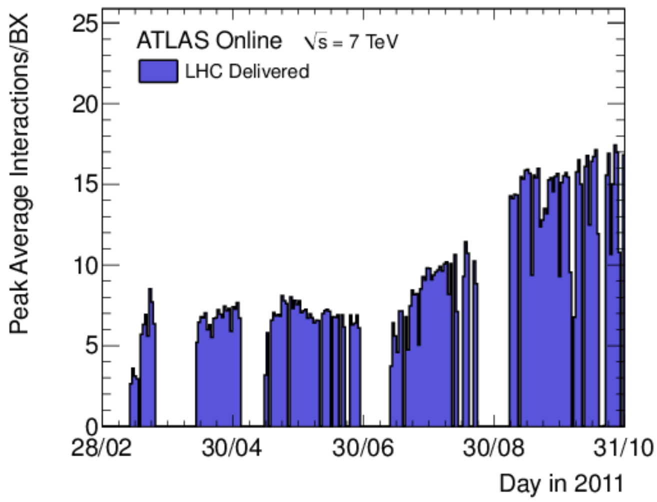

Figure 3.2: The maximum mean number of events per beam crossing versus day in 2011.

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

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 Σ_x and Σ_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 [63]. 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 *vdM* 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 erros evaluated see reference [62].

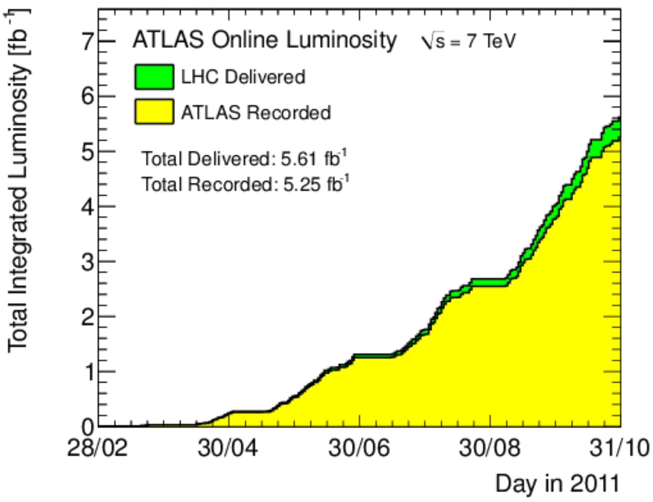

Figure 3.3: Total luminosity delivered by the LHC and recorded by ATLAS during the 2011 $sqrts = 7$ TeV proton-proton run

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 crossings 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 significatively 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 [64] 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 ϕ is measured as usual around the beam axis, and the polar angle θ is the angle from the beam axis. The pseudorapidity is defined as $\eta = lntan(\frac{\theta}{2})$, regions of low η are referred to as "central", and regions of high η are referred to as "forward". The transverse momentum p_T is defined in the x-y plane unless stated otherwise. The distance ΔR in the pseudorapidity-azimuthal angle space is defined as $DeltaR = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

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) systema 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$\sim$200 Hz. This requires a high rejection of minimum-bias processes while maintaining max-

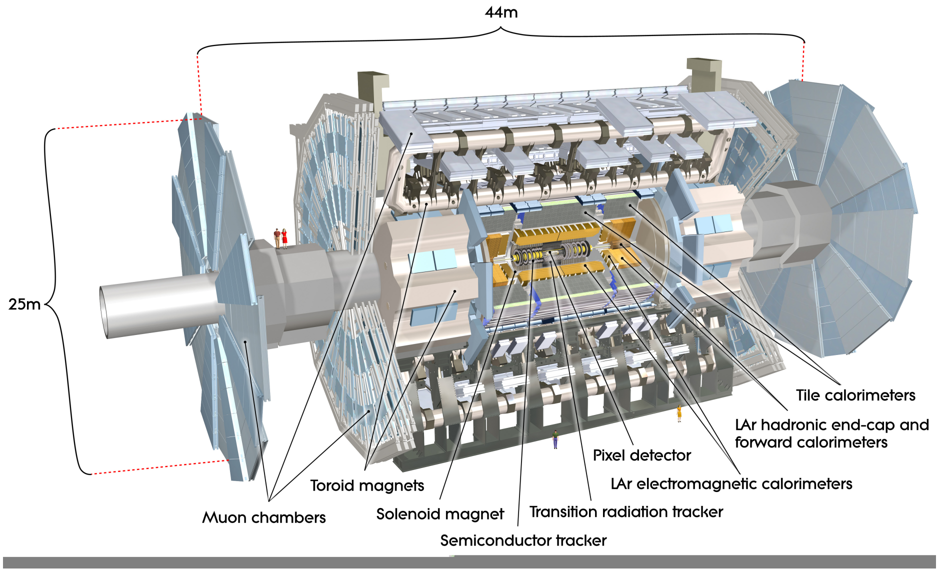

Figure 3.4: El detector de ATLAS

imum 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 subdetectors: 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 ϕ coverage.

The sensors which built this system register signals, referred to as "hits", in response to the passage of charged particles. The ID is inmersed 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 η coverage [65]. The transverse momentum resolution of $\sigma_{p_T}/p_T = 0.05$ [66] (upper bound) and a transverse impact parameter resolution of $\sim$20 μm for high momentum resolution particles in the central η region[67].

The pixel detector, SCT, and TRT sensors are arranged on concentric cylinders around the beam axis, known as barrel layers, and on disks perpindicular 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

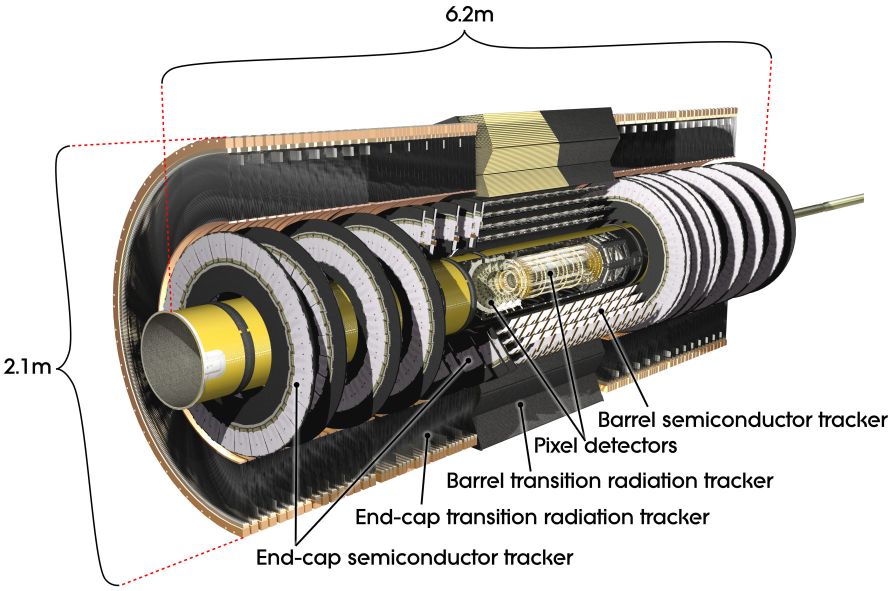

Figure 3.5: Layout of the ATLAS inner detector.

- $\phi \times z$ of 50 $\times$ 400 μm^2. The intrinsic accuracies in the barrel are 10 μm in r - ϕ and 115 μm 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 mm and radius of r ~150 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 mm and r ~515 mm, while the end-cap covers out to $z \sim \pm$2720 mm and r =560 mm. There are 15,912 SCT module sensors, each 12.8 cm long and approximately 285 μm 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 ϕ 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 μm. The intrinsic accuracies per module in the barrel are 17 μm in r - ϕ and 580 μm 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 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 silicom detectors and is comprised of up to 76 layers of longitudinal straw tubes in the barrel, extending to $z \sim \pm 710$ mm and $r \sim 1060$ mm, and 160 radial straw planes in each end-cap cylinders, reaching $z \sim \pm 2710$ mm and $r \sim 1000$ mm.

The TRT sensors are thin drift tubes consisting of cathode metal straws filled with an ionizing gas mixture of xenon, oxygen, and 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 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 ϕ 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.

The calorimeter detector consist of electromagnetic (EM) calorimeter and hadronic calorimeter components. The EM calorimter provides fine granularity measurements of electrons and photons. Each calorimeter is segmented both transverse to the particle direction, to give position information, and along the particle direction, to chart the development of the particle shower. This permits detailed mapping of EM and hadronic showers in the calorimeter, allowing for studies of the internal structure of hadronic jets and partially giving rise to the high resolution measurements of their energy.

The EM and hadronic 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.

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, and 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 mantained at a constant temperature of $\sim$88K. 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 all these components can be found in Fig. 3.6. Further specifications are given in the next sections.

Liquid argon EM calorimeter

The EM calorimeter uses lead as the absorber and liquid Argon as the active material. A photon traversing the absorber will interact with the heavy nucleus via Compton scattering or the photo-electric effect, producing low-

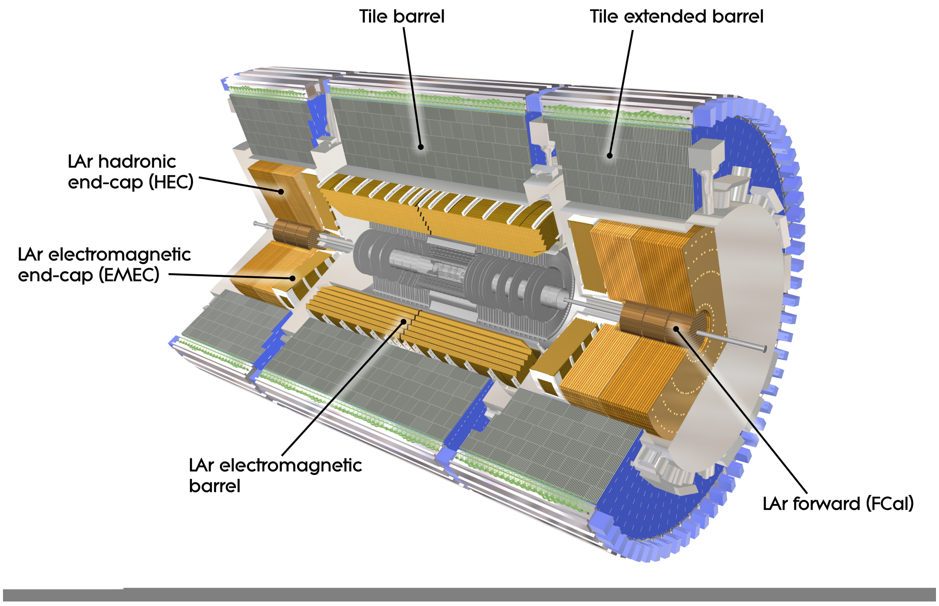

Figure 3.6: Layout of the ATLAS electromagnetic and hadronic calorimeter systems. The total length is $\sim$ 12 m, extending to a maximum radius of 4.25 m.

energy electrons, or pair production, producing electron/positron pairs. An electron or positron, in turn, can produce bremsstrahlung photons as it is deflected by the nuclei or produce more charged particles via ionization. Thus each incident photon, electron, or positron produces a shower of photons, electrons, and positrons that lose their energy through successive interactions in the absorber. The produced particles ionize the liquid argon, and the charge is collected by electrodes located in the liquid argon gap. These electrodes consist of three layers of copper sheets, the outer two kept at high-voltage potential and the inner one used to readout the signal.

To provide full coverage in ϕ without any cracks, an accordion-shaped absorber and electrode geometry is used, shown in Fig. 3.7. This design was chosen to ensure high azimuthal uniformity, a regular liquid argon ionization gap, and a constant sampling fraction within a given detector region. The figure highlights how this geometry is divided among rectangular cells in $\eta \times \phi$ space, the individual readout elements of varying size, finely segmented both laterally and longitudinally. Such fine segmentation $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ in the second layer of the EM barrel, for example permits a detailed mapping of the electromagnetic and hadronic showers.

The position resolution of the EM is driven by the readout geometry (rectangular cells). There are three layers of cells, segmented along the particle's direction of motion. The ϕ segmentation comes from grouping the accordion-shaped electrodes together into a common read out channel.

In the region $0 < |\eta| < 1.8$ the electromagnetic calorimeters are complemented by a "presampler" detector, an instrumented argon layer, which provides a measurement of the energy lost in the solenoid and the outer wall of the barrel cryostat.

The EMEC uses the same accordion geometry as the EMB, whereas the granularity is typically slightly larger than in the barrel.

The signal readout chain for the LAr calorimeter (indeed for all calorimeter systems) is divided into a fast analog readout for the trigger system and a slower digital readout used for more redefined trigger decisions and the offline reconstruction. However, regardless of the readout path, the signal is initiated within the active LAr medium. To minimize noise and increase speed the first level of readout is located on the detector (both for LAr and Tile calorimeter, see 3.2.2). The front-end electronics amplify and shape the signal. Shaping electronics induce a bipolar puse shape in the ionization signal. This shape is characterized by having both a positive and a negative component, which renders the integral of the signal exactly equal to zero.

The performance of the shaping electronics is critical for a correct energy calibration of the detector since the energy is primarily determined from the peak height of the pulse. In each calorimeter region, the overall pulse shape and duration are optimized to approximately cancel a constant injection of energy into the detector. The motivation for this approach is to effectively redefine the baseline of the energy measurement. In the high luminosity environment of the LHC, this reduces the sensitivity to the background from multiple *pp* interactions on average.

To translate these analog signals to digital signals that can be transmitted long distances to the next stage of the readout system, the pulse shape is measured over several 25 ns (nominal) time intervals, known as samples. The challenge of calorimeter calibration is to map these measured signals to the energy deposited in the active detector medium, known as the visible energy. This calibration is established using test-beam measurements of electrons in the EMB (REFERENCES) and EMEC calorimeters (REFERENCES).

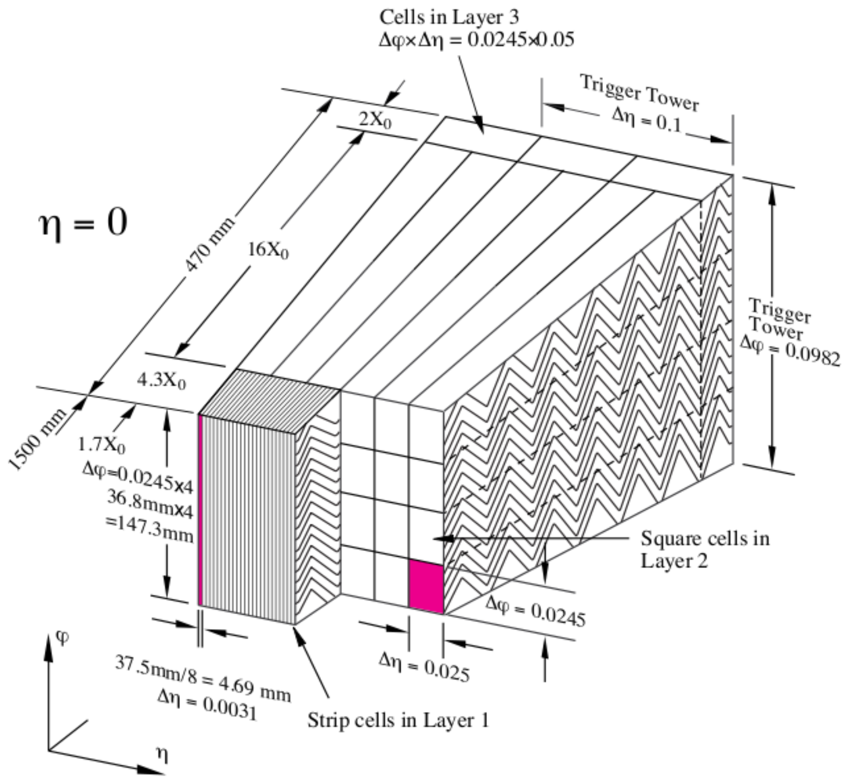

Figure 3.7: Cross section of the LAr barrel calorimeter where the different layers are visible. The granurality in η and ϕ of the cells of each of the three layers is also shown.

The hadronic calorimeter

Outside the EM calorimeter lies the system of hadronic calorimeters. The barrel portion, known as the Tile calorimeter, uses iron absorber slabs interspersed with scintillating tiles. The Tile calorimeter is most notable for its depth of 7.4 radiation lengths (λ[1]). The hadronic end-cap and the forward calorimeter, which need to absorb the more energetic particles that are produced at large $|\eta|$, are made of copper and tungsten abosrbers, respectively, with liquid argon as the active material.

The tile calorimeter is composed of 3 mm thick scintillating tiles, arranged to lie parallel to the incoming particle direction, interleaved with 14 mm thick iron plates. It is divided into the barrel calorimeter, covering $|\eta| < 1.0$, and two extended barrel calorimeters, covering $0.8 < |\eta| < 1.7$. Each tile is read out by two wavelength-shifting fibers, which convert the scintillator signal to visible light. The readout fibers of several tiles are grouped to a single photomultiplier tube forming cells in $\eta \times \phi$ space. As in the EM calorimeter, these cells are segmented into three layers, the first two of size $\Delta\eta = 0.1$ and $\Delta\phi = 0.1$ and the last of size $\Delta\eta = 0.2$ and $\Delta\phi = 0.1$. Towers to provide information to the trigger systema are formed from 0.1×0.1 grouping of all three layers.

The HEC uses the LAr active readout design due to the higher radiation tolerance required for the forward regions. Although housed in the same cryostat as the accordion geometry EMEC, the HEC implements a flat-plate design.

The forward calorimeter extends to cover the region $3.1 < |\eta| < 4.9$. Since

[1] To quantify the amount of material needed to capture a particles's energy, the unit of an interaction length, which is the distance over which a high energy charged particle loses $1 - \frac{1}{e} \sim 63\%$ of its energy , is commonly used.

it is the only calorimeter that covers this very forward region, it must provide both electromagnetic and hadronic measurements. In addition, the high particle fluxes in this region necessitate a finely granulated design. The FCal is approximately 10 interaction lenghts deep, and consists of three modules in each end-cap: the first, made of copper, is optimised for electromagnetic measurements, while the other two, made of tungsten, measure predominantly the energy of hadronic interactions.

The hadronic calorimeters are calibrated using muons in test-beam experiments (REFERENCIAS) and those muons produced by cosmic-rays in situ (REFERENCIA). The invariant mass of the Z boson in $Z \to ee$ events measured in-situ in the 2010 pp collisions is then used to adjust the calibration derived from test-beams and cosmic-muons.

3.2.3 The Muon System

The muon system gives the ATLAS detector its overall shape and imposing nature, as depicted in Fig. 3.8. Muons have much smaller cross section to interact in material than electrons and hadrons, for this, they do not deposit all their energy in the calorimeters. The muon spectrometer is designed to detect muons within $|\eta| < 2.7$. Because many new physics signatures involve high-momentum muons, the system is also required to provide trigger signals based on the particle p_T for $|\eta| < 2.4$.

To provide a momentum measurement, the muons trajectorires are bent in a toroidal magnetic field. This field is provided by one large barrel toroid and two large end-cap toroids, each toroid consisting of eight coils arranged symmetrically around the beam axis. The toroid system produces a magnetic field that is typically oriented in the ϕ direction and that is measured with over 1800 Hall sensors placed throught the magnets. Under the influence of this field, muons are deflected in the $r-z$ plane and the transverse momentum of the muons is given then by the radius of curvature of the tracks. Since the highly-energetic muons bend very little even in this high magnetic field, the muon systema is the largest of all the ATLAS sub-detectors, covering a radius from $\sim$4.5 m to $\sim$12.5 m.

Four primary subsystems comprise the integrated muon spectrometer: monitored drift tubes (MDT), cathode strip chambers (CSC, which are multiwired proportional chambers with cathodes segmented into strips), resistive plate chambers (RPC) and thing gap chambers (TGC). The MDT and CSC subsytems are primarily designed for precision measurements of muon tracks, width the MDT system providing coverage for the more central region ($|\eta|$ ¡ 2.7, with full coverage only in $|\eta| < 2.0$), whereas the CSC is located in the more forward region ($2.0 < |\eta| < 2.7$) due to its ability to cope with higher background rates. The RPC and TGC muon subsystems are desinged to provide fast, robust readout for use in the trigger and data acquisition system.

A description of the subsystems can be found elsewhere (REFERENCIAS!!!).

3.2.4 Forward detectors

Three smaller detector systems cover the ATLAS forward region. The main function of the first two systems is to determine the luminosity delivered to ATLAS. At $\pm$17 m from the interaction point lies LUCID (LUminosity measurement using Cerenkov Integrating Detector). The principle of LUCID is to detect inelastic $p-p$ scattering in the forward region, exploiting the fact that the number of particles detectted is proportional to the total, both primary and pileup, interactins in a bunch-crossing. LUCID thus provides a

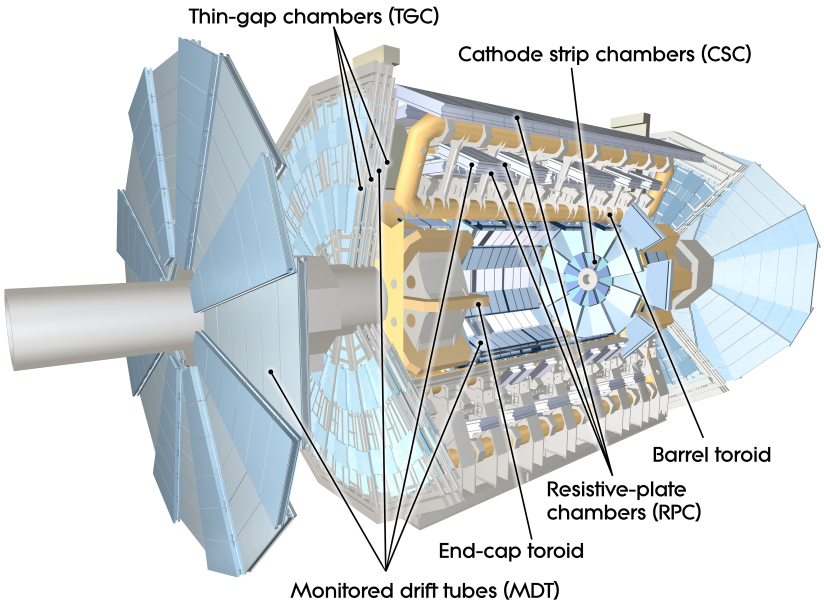

Figure 3.8: Muon Chamber

relative luminosity measurement, in which the detected number of particles must be translated to the total number of proton-proton interactions via calibration runs. The second detector is ALFA (Absolute Luminosity For ATLAS). Located at ±240 m, it consists of scintillating fibre trackers located inside Roman pots which are designed to approach as close as 1 mm to the beam. The third system is Zero-Degree Calorimeter (ZDC), which place a role in determining the centrality of heavy-collisions.

3.2.5 Trigger and Data Adquisition

At design luminoisty, the LHC will deliver approximately 40 million collision events every second. With an average ATLAS event size of ∼1.5 MB, this is far more information than can be saved into the finite data storage resources available. The goal of the trigger system is to move interesting physics events to permanent storage, while rejecting the vast majority of other events.

The online selection is done in three stages: the Level 1 (L1), Level 2 (L2) and Event Filter (EF) stages. Each trigger level refines the decisions made at the previous level and, where necessary, applies additional selection criteria. The data acquisition (DAQ) system receives and buffers the event data from the detector-specific readout electronics, at the L1 trigger accept rate, over 1600 point-to-point readout links. The L1 trigger uses a limited amount of the total detector information (only from the calorimeter and the muon systems) using only simple hardware based algorithms to make a decision in less than 2.5 μs, reducing the rate to about 75 kHz. The L2 and EF, collectively referred to as the High Level Trigger (HLT), are based on fast software algorithms running on large farms of commercial processors. The L2 is the first stage of the ATLAS DAQ system that has access to data from the ID and is capable of doing partial reconstruction of events up to the L1 accept rate. L2 trigger is designed to reduce the rate to approximately 3.5 kHz, with an event processing time of about 40 ms, averaged over all events. The EF reduces the rate to roughtly 200 Hz. Its selections are implemented using offline analysis procedures within an average event processing time of the order of 4 s.

The L1 trigger is desinged to accept high-p_T muons, electrons, photons, jets, and taus, as well as events with large missing transverse energy or sum energy. It uses signals from the TGCs and RPCs from muon triggers and reduced granularity calorimeter information for electron, photon, jet, tau and total energy triggers. The calorimeter trigger system, which maintains a fast readout independent from the remainder of the calorimeter is known as the Level-1 Calorimeter. At this level coarse calorimeter information is available in the form of jet elements with $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ for $|\eta| < 3.2$. Jets are reconstructed using a square sliding window algorithm. In addition to coarse jets, the total transverse energy is also measured at the L1. The region of the detector corresponding to the location where the L1 thresholds were passed − so called "region of interes" (RoI) − are then delivered to the L2 sofware algorithms.

The L2 trigger applies additional energy thresholds and multiplicity requirements using the RoI around triggered L1 objects. For example, the L2 jet trigger retrieves the data from cells surrounding the L1 RoI and constructs jets using a simplified cone jet algorithm.

The next step and last stage in the trigger chain is the EF, which receives events that have been selected by the L2 triggers and processes the entire event with the full detector granularity instead of only a restricted region.

The monitoring infrastructure of the HLT supports the real-time accumulation of histograms, and their aggregation across the farm, so that parameters can be extracted from cumulative distributions that contain events from all processor nodes. Beam parameters determined from those live histograms are transmitted online to the LHC and are also available to fed back into the HLT itself for use by its own trigger algorithms that depend on the precise knowledge of the luminous regino (such as b-jet tagging).

3.2.6 ATLAS Performance and Data quality

The ATLAS detector has been operational for a number of years collecting large amounts of data. Before the start-up of the LHC, measuring muons from cosmic rays, which were used to test, understand and align the detector. In 2010 and 2011 ATLAS recorded over 5.2fb^{-1} of collision data. Fig. 3.3 presented the luminosity delivered by the LHC in 2011 as well as the recorded luminosity by the detector, showing a good performance of the

ATLAS Experiment.

The fraction of time that each subdetector system was operational during data-taking is shown in Table 3.2.

The Data Quality (DQ) selection within ATLAS is based on the inspection of a standard set of distributions that lead to a data quality assesment which is encoded in so-called DQ flags. DQ flags are issued for each detector, usually segmented in subdetectors like barrel, end-caps and forward. DQ flags are also issued for trigger slices and for each physics object reconstruction. In this way, the state of the ATLAS detector from hardware to physics object reconstruction is expressed through DQ flags, which are saved per luminosity block. A luminosity block is a time interval of typically two minutes.

The DQ information is used in analyses through dedicated lists of good runs/luminosity blocks. Good run lists are formed by DQ selection criteria in addition to other criteria, such as run range, magnetic field configuration and beam energy. A complete list of valid physics runs and luminosity blocks is used in each analysis.

Detector component	operational
Inner Detector	
Pixel	≈96.4%
SCT	≈99.2%
TRT	≈97.5%
Calorimeter	
EM	≈99.8%
Tile	≈96.2%
Hadronic, end-cap	≈99.6%
Forward calorimeter	≈99.8%
Muon Spectrometer	
MDT	≈99.7%
CSC	≈97.7%
RPC	≈97.0%
TGC	≈97.9%

Table 3.2: The approximate fraction of time that each individual subdetector system was operational during data-taking.

Chapter 4

Jet reconstruction and b-Tagging

The reconstruction of the key objects for b-tagging purposes, namely the tracks, the primary vertex and jets is briefly described in the following.

4.1 Jets reconstruction and calibration

Jets are reconstructed using the anti-k_t jet algorithm [37] with a distance parameter $R = 0.4$, using calorimeter topological clusters [68] 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 [69].

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 [70]. 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 estabished. 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 [71].

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 [72]. 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 [73].

The ability to identify jets containing b-hadrons is important for the high-p_T physics program of a general-purpose experiment at the LHC such as ATLAS. Two robust b-tagging algorithms taking advantage of the impact parameter of tracks (JetProb) or reconstructing secondary vertices (SV0) have been quickly commisssioned [74][75] and used for several analyses of the 2012 and 2011 data (REFERENCIAS). Building on this success, more advanced b-tagging algorithms have been commissioned with the 2011 data. All these algorithms are based on Monte Carlo predictions for the signal (b-jet) or background (light- or in some cases c-jet) hypotheses.

The b-tagging performance relies criticaly on the accurate reconstruction of the charged tracks in the ATLAS Inner Detector. The innermost part, the pixel detector, has an intrinsic measurement accuracy of around 10 μm in the transverse plane, and 115 μm along the beam axis (z). For a central track with p_T =5 GeV, which is typical for b-tagging, the transverse momentum resolution is around 75 MeV and the transverse impact parameter resolutioni is about 35 μm.

4.2.1 Primary vertex

The knowledge of the position of the primary interaction point (primary vertex) of the proton-proton collision is important for b-tagging since it defines the reference point with respect to which impact parameters and vertex displacements are measured.

See primary vertex recontruction in [76].

4.2.2 Pile-up

Out-of-time pile-up events (pp collisions from neighboring bunches in the same train) also generate calorimeter activity and consequently extra jets. However, given the time resolution of the Inner Detector, and since the b-tagging algorithms reject jets with no track associtaed to them, the contribution of the out-of-time pile-up for this analysis is expected to be negligible.

4.2.3 Tracks selection and properties

Track quality cuts

The track selection for b-tagging is designed to select well-measured tracks rejecting fake tracks and tracks from long-lived particles (K_s, Λ, and other hyperon decays, generically referred to as V^0 decays) and material description.

The tracks of charged particles with a pseudorapidity $|\eta| < 2.5$ are reconstructed in the the Inner Detector. It is composed of a barrel, consisting of 3 Pixel layers, 4 double layers of single-sided silicon strip sensors, and 73 layers of Transition Radiation Tracker straws concentric with the beam, plus a system of disks on each end of the barrel, occupying in total a cylindrical volume around the interaction point of radius of 1.15 m and length of 7.024 m. The Pixel detector's innermost layer is located at a radius of 5 cm from the beam axis, has a position resolution of approximately 10 μm in the $r - \phi$ plane and 115 μm along the beam axis (z).

Track association to jets

The actual tagging is performed on the sub-set of tracks in the event that are associated to the jets. Tracks are associtaed to the jets with a spatial matching in $\Delta R_{(jet,track)}$. The association cut ΔR is varied as a function of the jet p_T in order to have a smaller cone for high-*pt* jets which are more collimated.

Impact parameters

The most critical track parameters for b-tagging are the transverse and longitudinal impact parameters. The transverse parameter d_0 is the distance of closest approach of the track to the primary vertex point in the $r\phi$ projection. The z coordinate of the track at this point of closest approach is referred to as z_0. It is often called the longitudinal impact parameter[1]. On the basis that the decay point of the b-hadron must lie along its flight path, the impact parameter is signed to further discriminate the tracks from b-hadron decays from tracks originating from the primary vertex. The sign is positive if the track extrapolation crosses the jet direction in front of the primary vertex, and negative otherwise. Therefore, tracks from b/c hadron decays tend to have positive sign.

The significance, which gives more weight to tracks measured precisely, is the main ingredient of the tagging algorithms based on impact parameters.

4.2.4 b-tagging algorithms

[1]Stricktly speaking the impact parameter is $|z_0|sin\theta$, where θ is the polar angle of the track.

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 [77] 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 [29]. Multiple parton interactions [30], as well as fragmentation and hadronisation based on the Lund string model [78] 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 [?] software toolkit within the ATLAS simulation framework [79] 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 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 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 [80] and with PYTHIA8 using the Perugia tune [32]. The former is based on the event generator HERWIG, but redesigned in the C++ programming 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 [81]. 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 ΔR 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 $\Delta R(jet^{\mathrm{reco}}, \mathrm{track}) \leq 0.4$ around the jet axis were considered. Cuts on $p_{\mathrm{T}}^{\mathrm{trk}} > 1.0$ GeV and the χ^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 θ 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 [41], 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, τ_1, τ_2, and ΔR 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. τ_1 and ΔR between the k_T axes are nearly linearly related, which is expected if there are two hard lobes of energy. But τ_2 is almost independent of ΔR between the k_T axes, meaning that regardless of where the axes are, the energy is uniformly distributed around them.

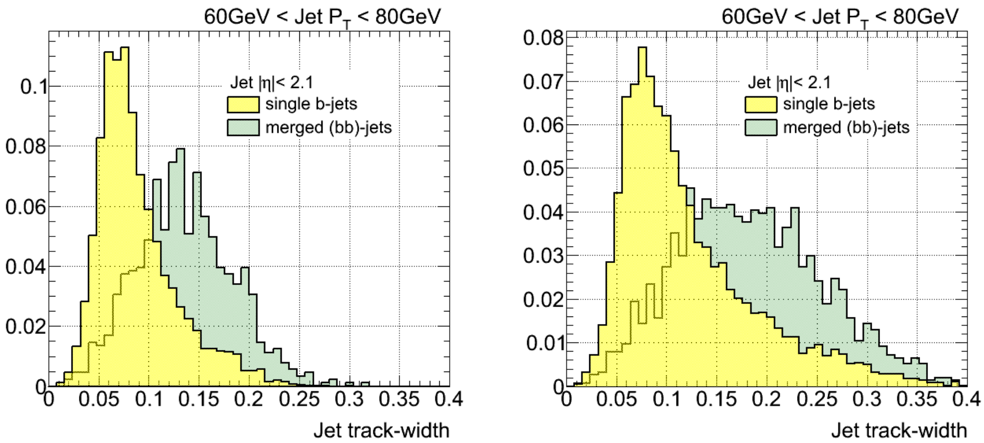

Figure 5.1: Distribution of track-jet width in anti-k_T 0.4 (left) and 0.6 (right) jets, for single and merged b-jets between 60 GeV to 80 GeV. Jets were built using all stable particles in the simulation.

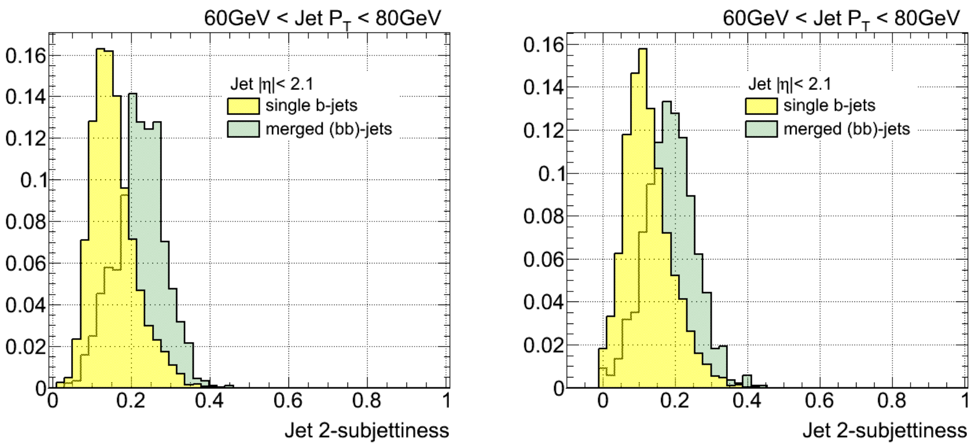

Figure 5.2: Distribution of τ_2 in anti-k_T 0.4 jets, for single and merged b-jets between 60 GeV to 80 GeV. Jets were built using all stable particles (left) and charged particles only (right).

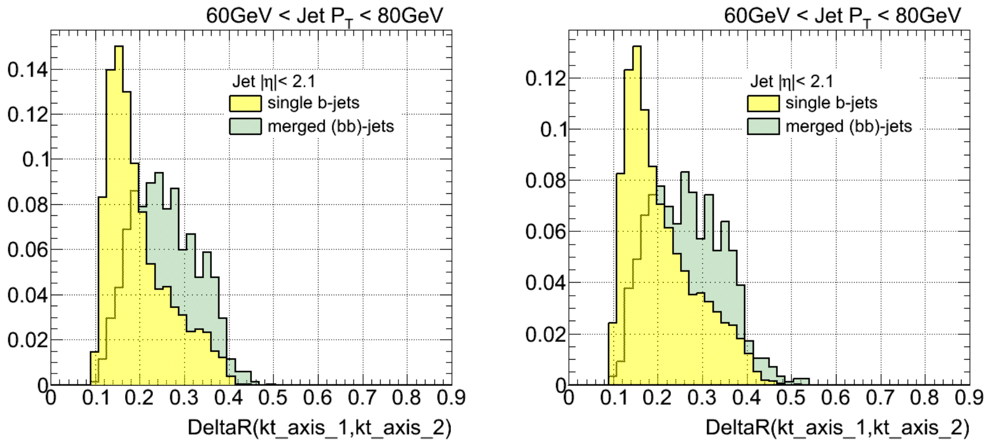

Figure 5.3: Distribution of ΔR between the axes of two k_T subjets in anti-k_T 0.4 jets, for single and merged b-jets between 60 GeV to 80 GeV. Jets were built using 0.1 k_t jets from all stable particles (left) and charged particles only (right).

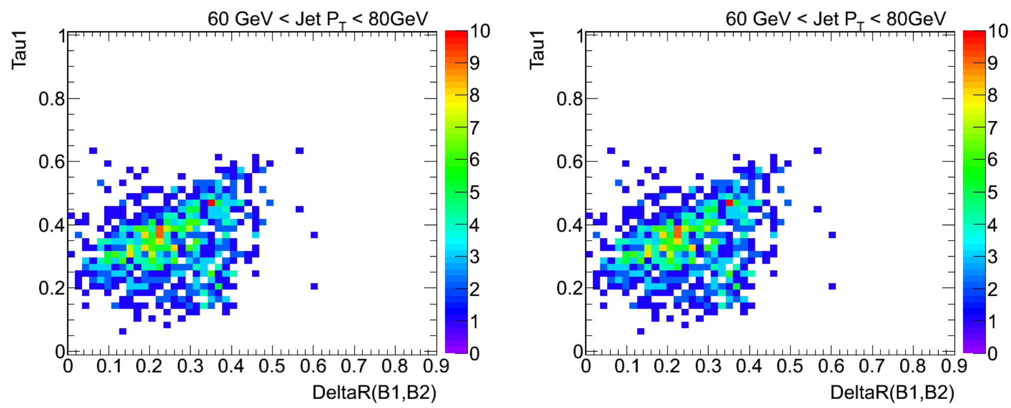

Figure 5.4: Correlation between τ_1 (left) and τ_2 (right) and the ΔR between the B-hadrons in merged anti-k_T 0.4 jets between 60 GeV to 80 GeV. Jets were built using all stable particles.

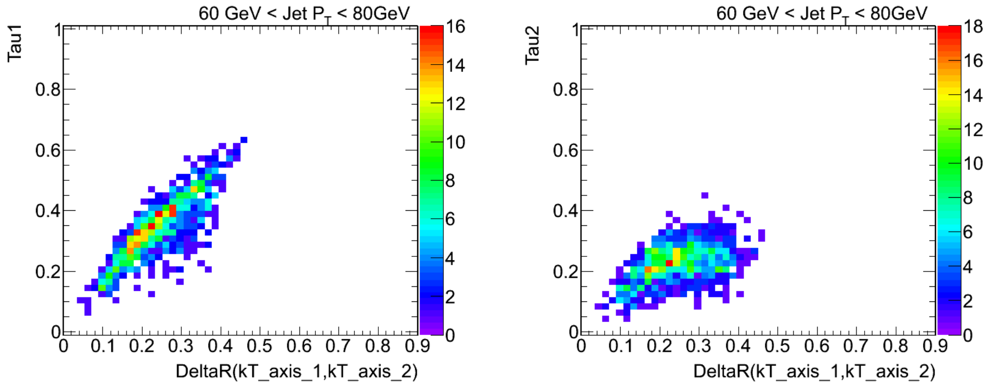

Figure 5.5: Correlation between τ_1 (left) and τ_2 (right) and the ΔR between the k_T subjets in merged anti-k_T 0.4 jets between 60 GeV to 80 GeV. Jets were built using all stable particles.

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 ΔR between pairs of constituents

- ΔR between 2 k_t subjets within the b-jet
- τ_2: 2-subjettiness
- τ_2/τ_1
- ΔR 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 ther 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 ΔR distance between the associated constituent ("*const*") and the jet axis:

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

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

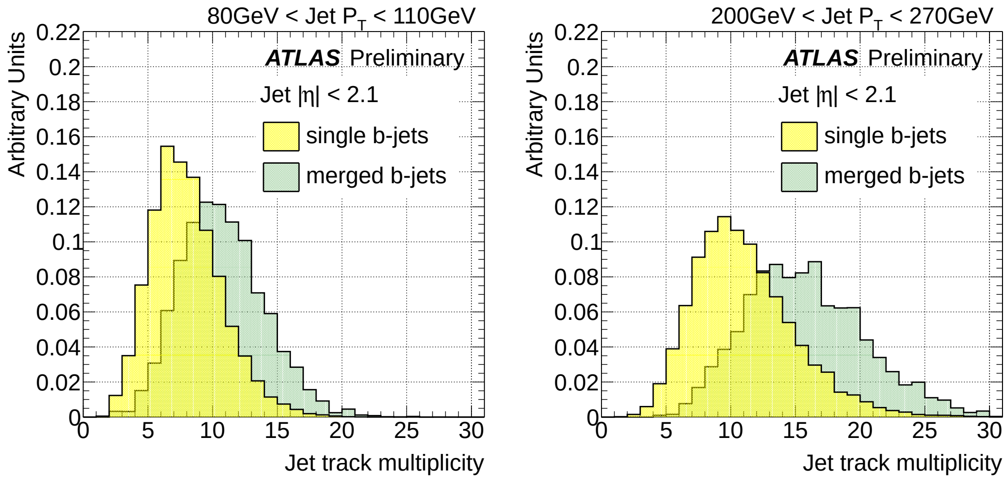

Figure 5.6: Distribution of the track multiplicity in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

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.

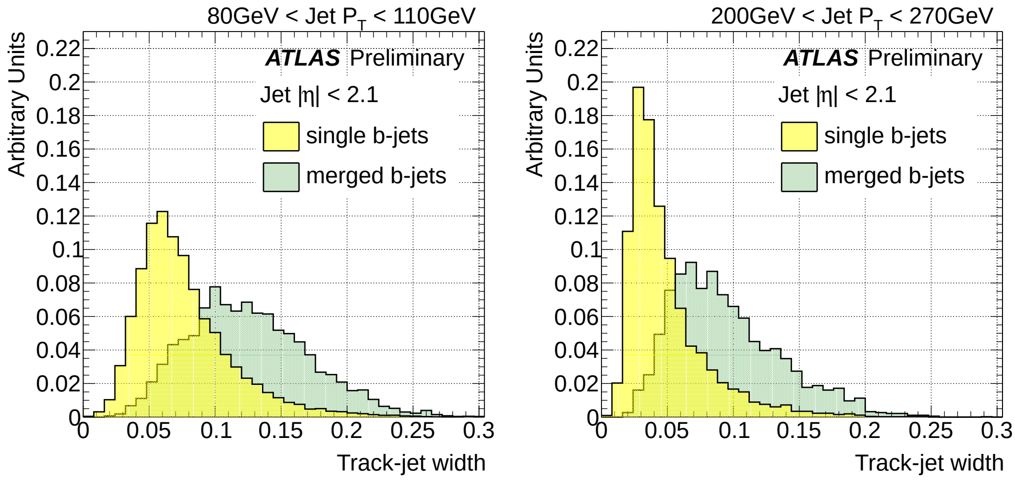

Figure 5.7: Distribution of track-jet width in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

III. Maximum ΔR between track pairs

Figure 5.11 shows the distribution of the maximum ΔR between track pairs in the jets (Max$\{\Delta R(trk, trk)\}$). 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 Max$\{\Delta R(trk, trk)\}$ as it is not an infrared safe observable and is sensitive to soft tracks originating from pile-up.

IV. ΔR between the axes of two k_t subjets

The distribution of the ΔR 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 [40] is applied to all the tracks associated to

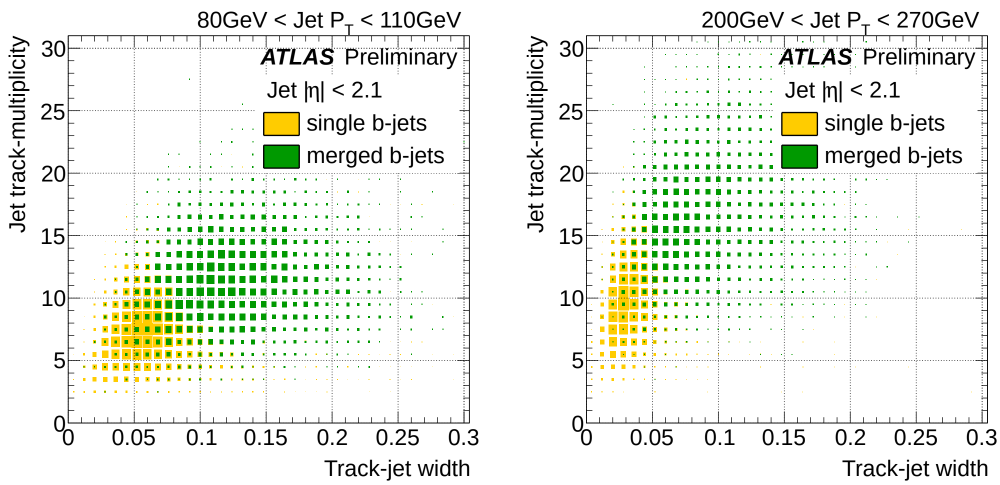

Figure 5.8: Correlation between jet track multiplicity and track-jet width for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

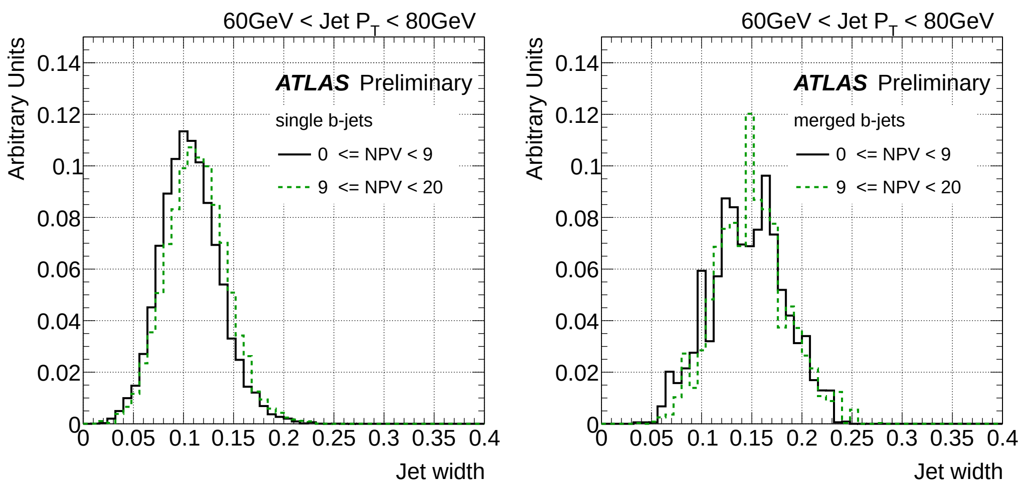

Figure 5.9: Distribution of calorimeter jet width (using topological clusters) for single (left) and merged (right) b-jets in two bins of Number of Primary Vertices for jets between 60 GeV to 80 GeV.

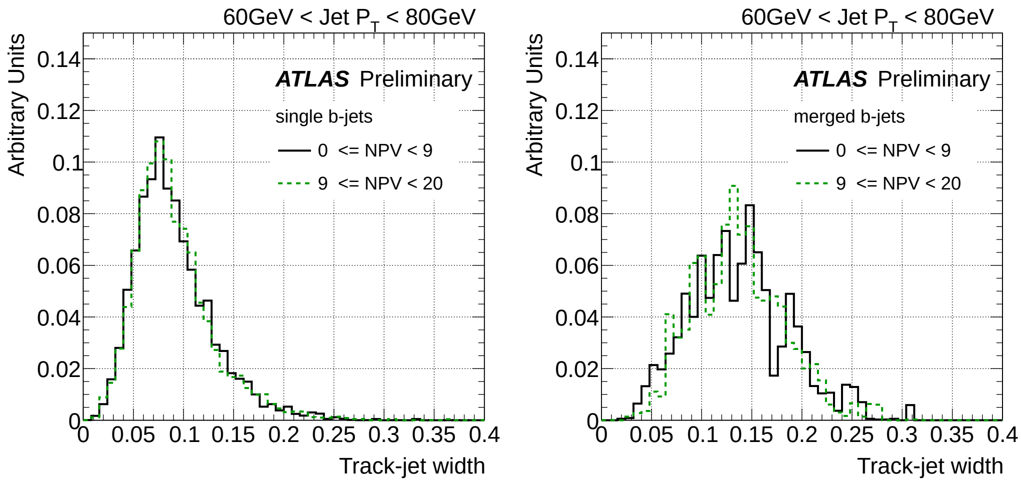

Figure 5.10: Distribution of track-jet width for single (left) and merged (right) b-jets in two bins of Number of Primary Vertices for jets between 60 GeV to 80 GeV.

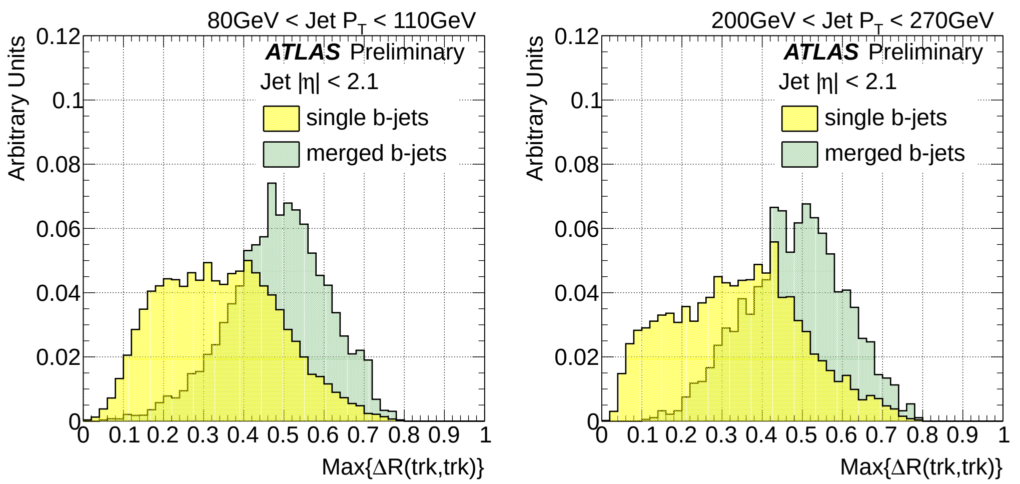

Figure 5.11: Distribution of the maximum ΔR between pairs of tracks in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

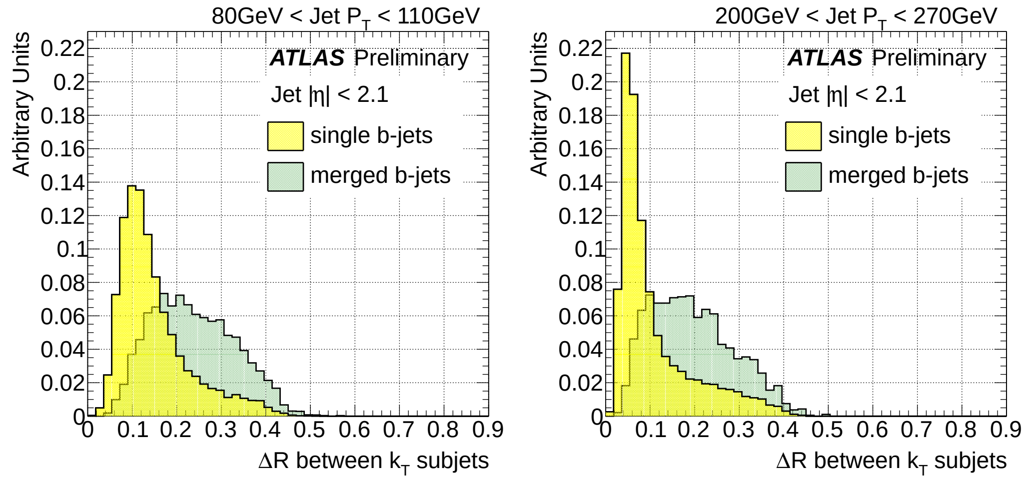

Figure 5.12: Distribution of the ΔR between the axes of the two k_t subjets in the jet for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

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. [56], 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:

$$\tau_N = \frac{1}{\sum_k p_{Tk} R_0} \sum_k p_{Tk} \min\{\Delta R_{S_1,k},\ \Delta R_{S_2,k}, ...,\ \Delta R_{S_N,k}\} \tag{5.2}$$

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 τ_2 value than a jet with tracks uniformly distributed in $\eta - \phi$ space.

Plots of τ_2 are shown in Fig. 5.13. In spite of its expected 2-prong substructure, merged b-jets have higher values of τ_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 τ_2. This suggests to switch from an absolute to a width-normalized τ_2. Fig. 5.15 thus shows the distributions of τ_2/τ_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

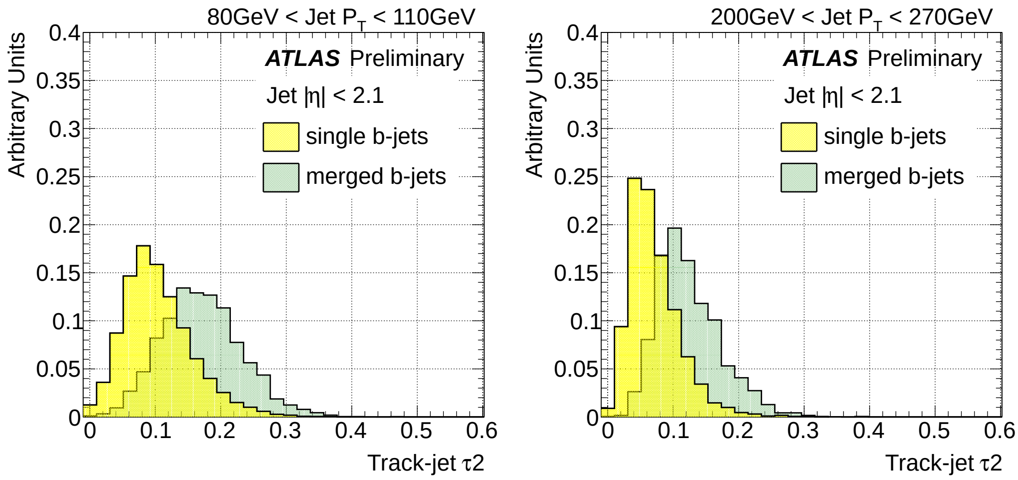

Figure 5.13: Distribution of τ_2 in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

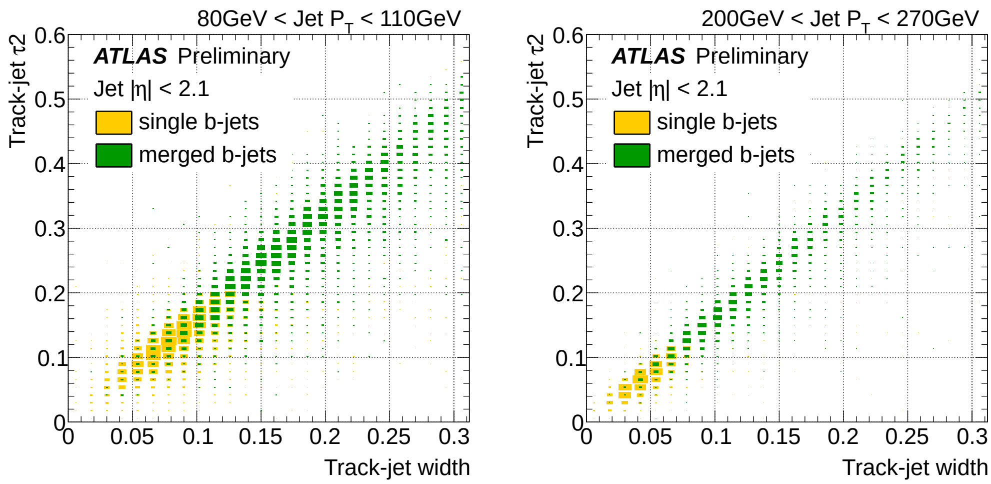

Figure 5.14: Correlation between τ_2 and track-jet width for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

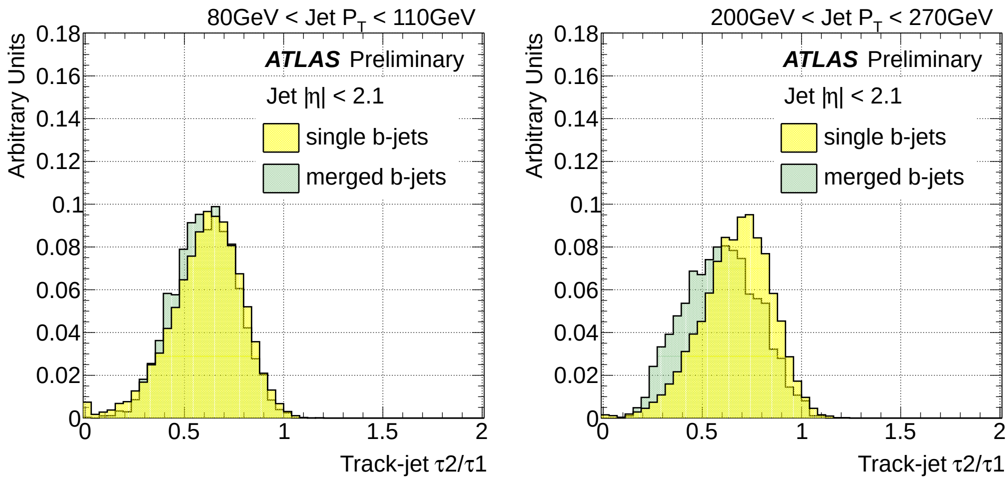

Figure 5.15: Distribution of τ_2/τ_1 in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

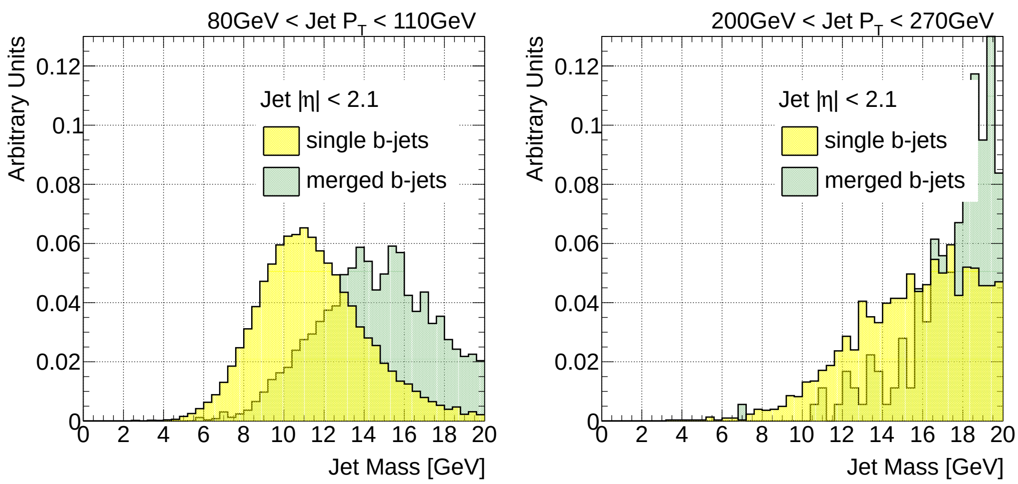

Figure 5.16: Distribution of jet mass in GeV for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

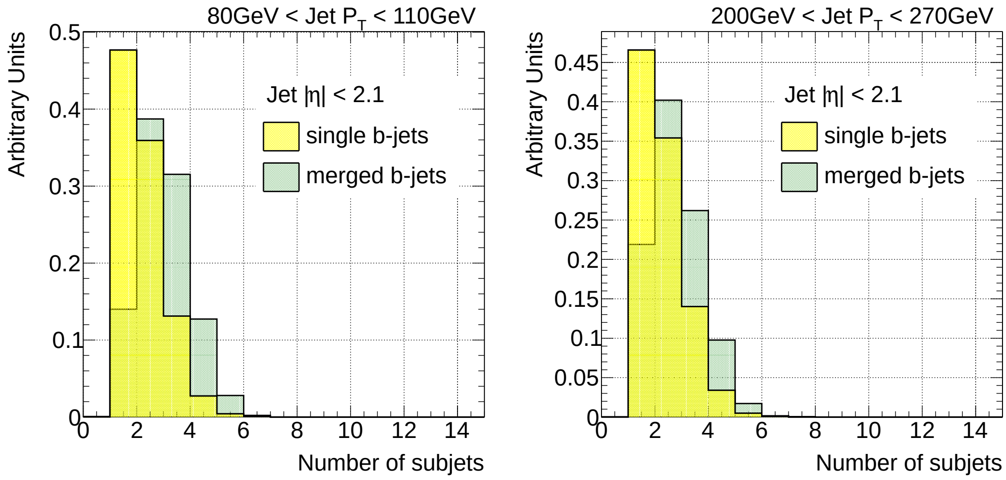

Figure 5.17: Distribution of the number of k_t sub-track-jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

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

VIII. ΔR between leading constituents

Figure 5.18 shows the distribution of the number ΔR 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.

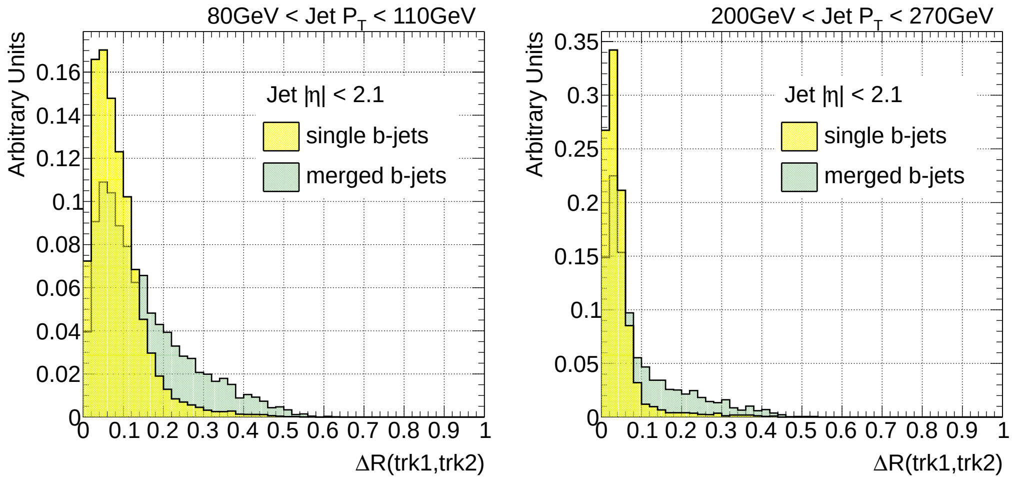

Figure 5.18: Distribution of ΔR between leading tracks for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

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 discrimation 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 τ_2, ΔR 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-

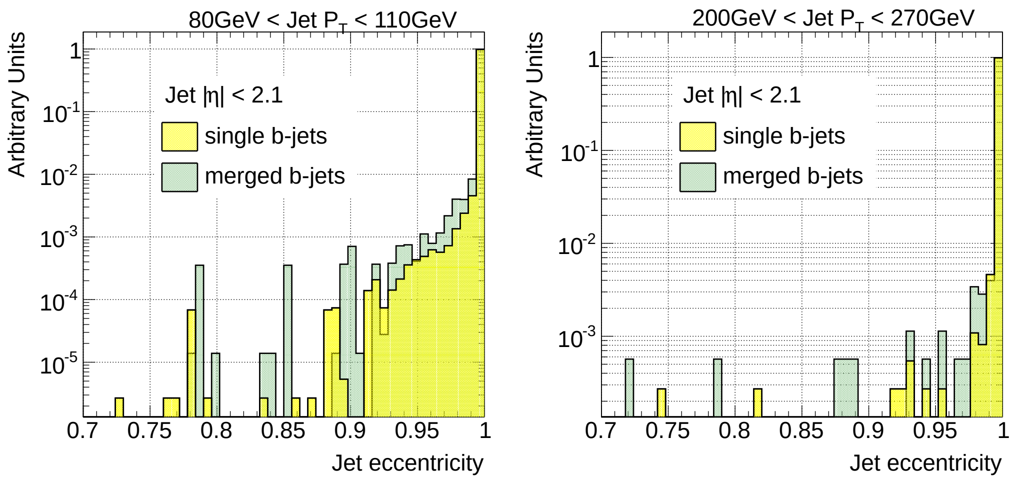

Figure 5.19: Distribution of the jet eccentricity for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

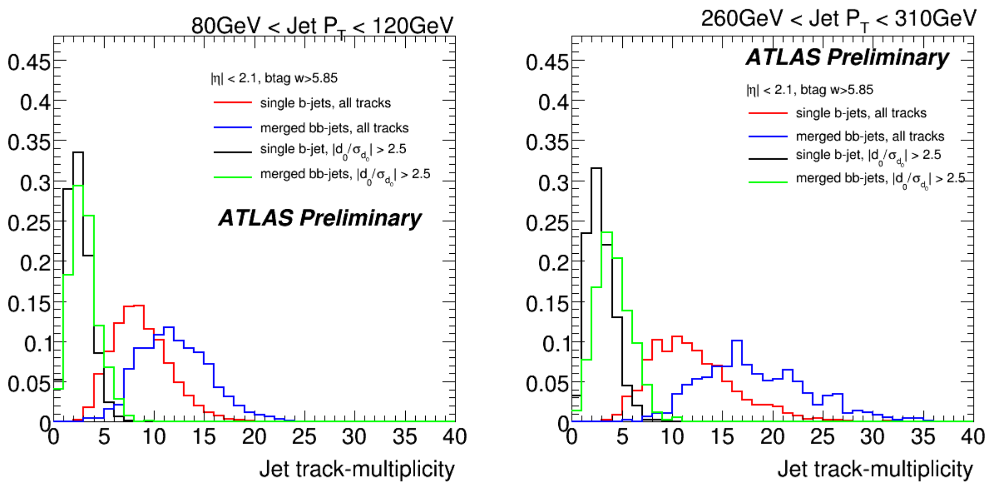

Figure 5.20: Distribution of the jet track multiplicity single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right), for all and displaced tracks only.

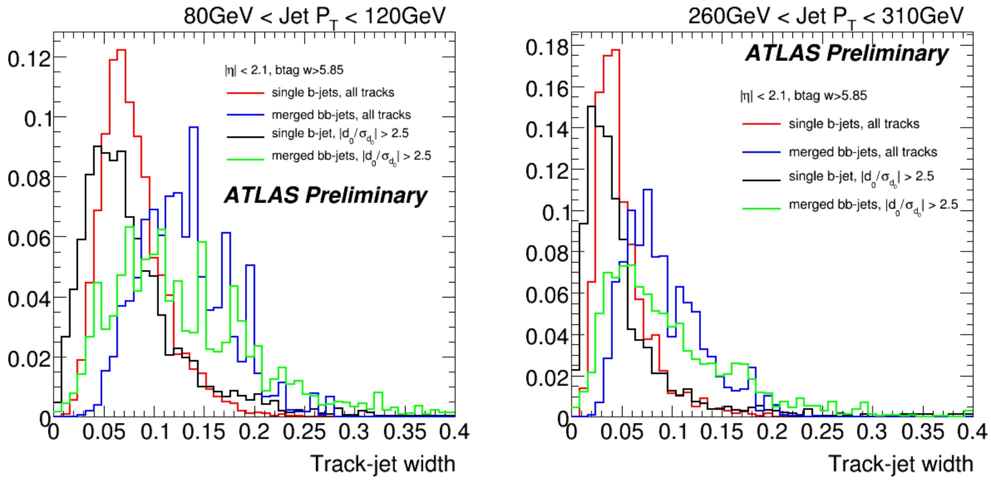

Figure 5.21: Distribution of the track-jet width for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right), for all and displaced tracks only.

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 ΔR 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-

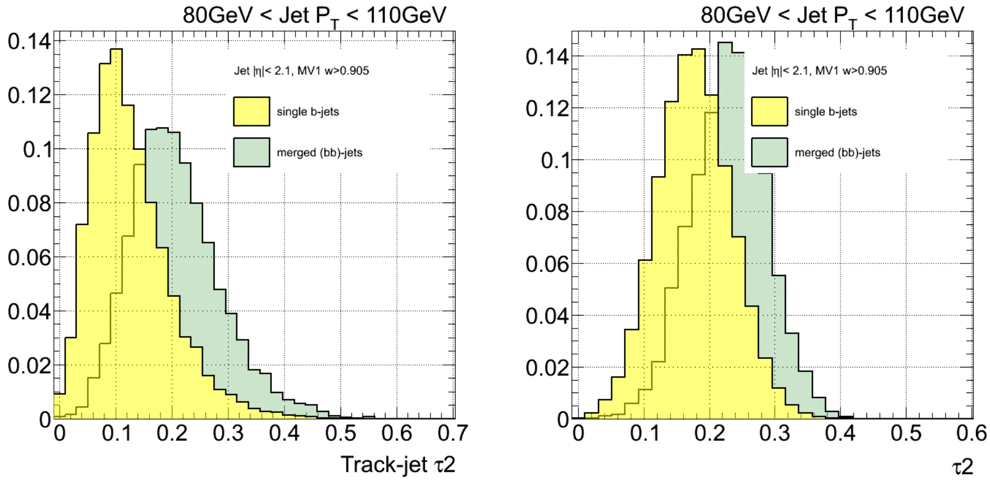

Figure 5.22: Distribution of τ_2 for single and merged b-jets between 80 GeV to 110 GeV in anti-k_T 0.6 jets using track constituents (left) and anti-k_T 0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

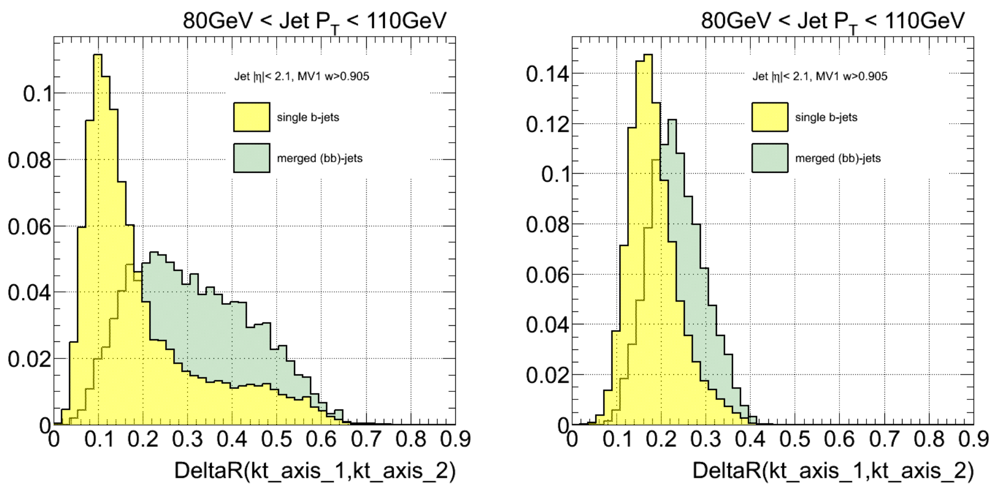

Figure 5.23: Distribution of ΔR between k_T subjets for single and merged b-jets between 80 GeV to 110 GeV in anti-k_T 0.6 jets using track constituents (left) and anti-k_T 0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

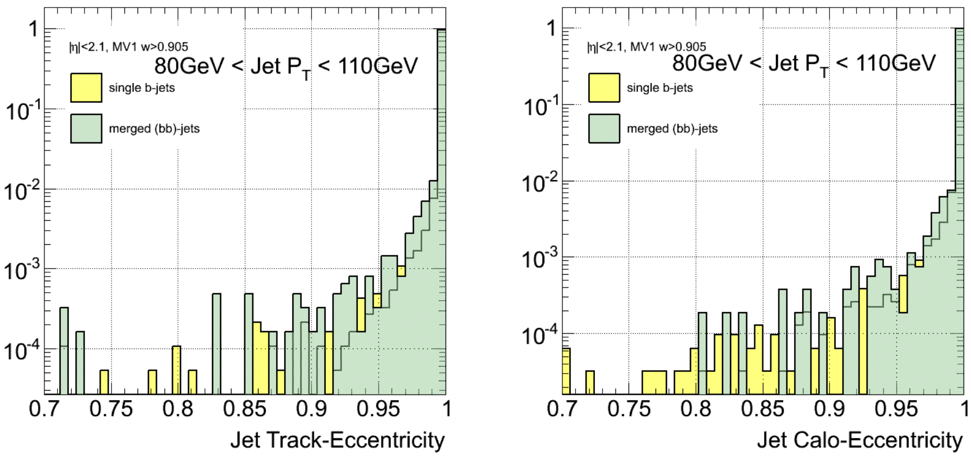

Figure 5.24: Distribution of the jet eccentricity for single and merged b-jets between 80 GeV to 110 GeV in anti-k_T 0.6 jets using track constituents (left) and anti-k_T 0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

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.

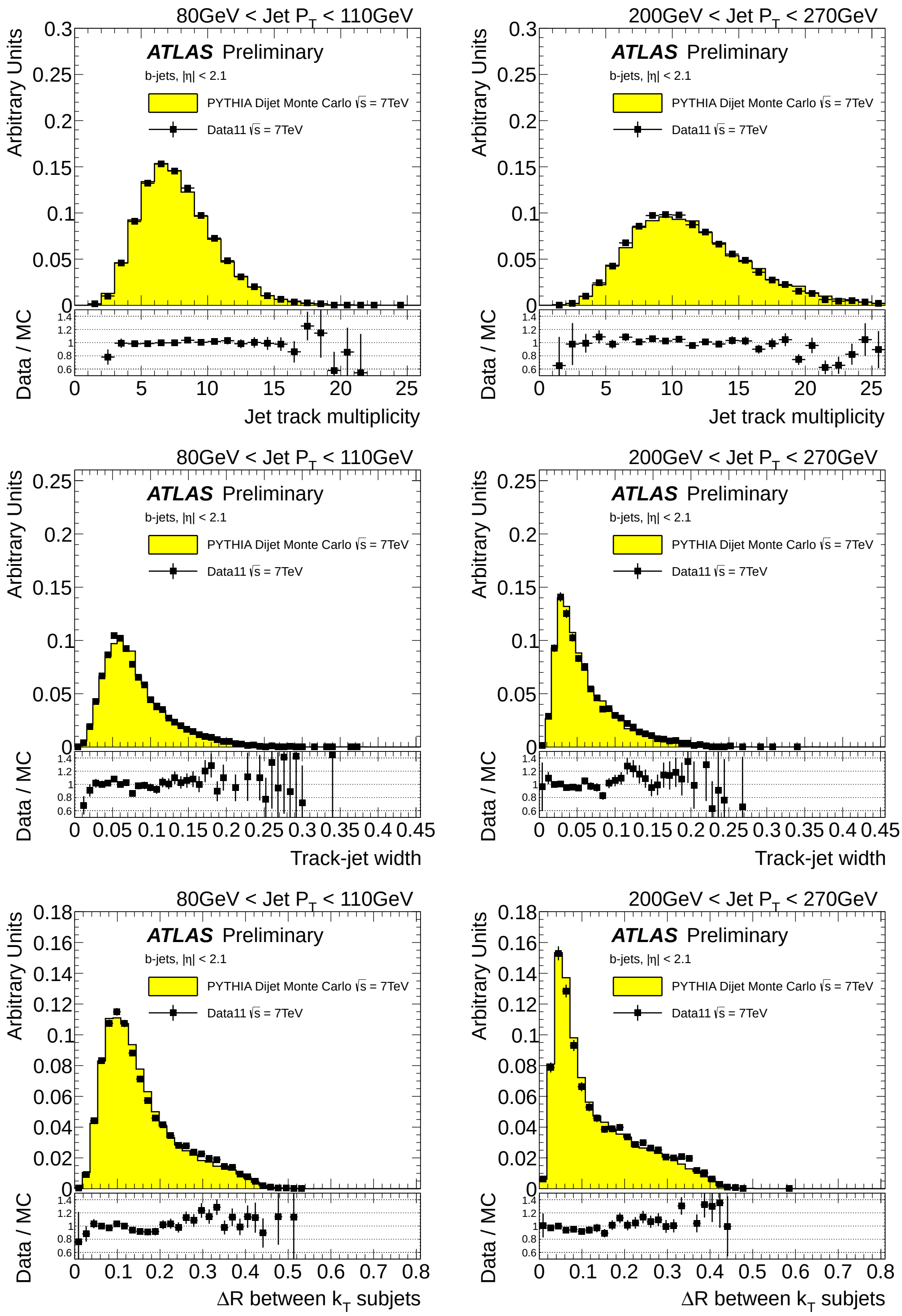

Figure 5.25: Distribution of three tracking variables in 2 different jet p_T bins, for experimental data collected by ATLAS during 2011 (solid black points), and simulated data (filled histograms). The ratio data over simulation is shown at the bottom of each plot.

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.

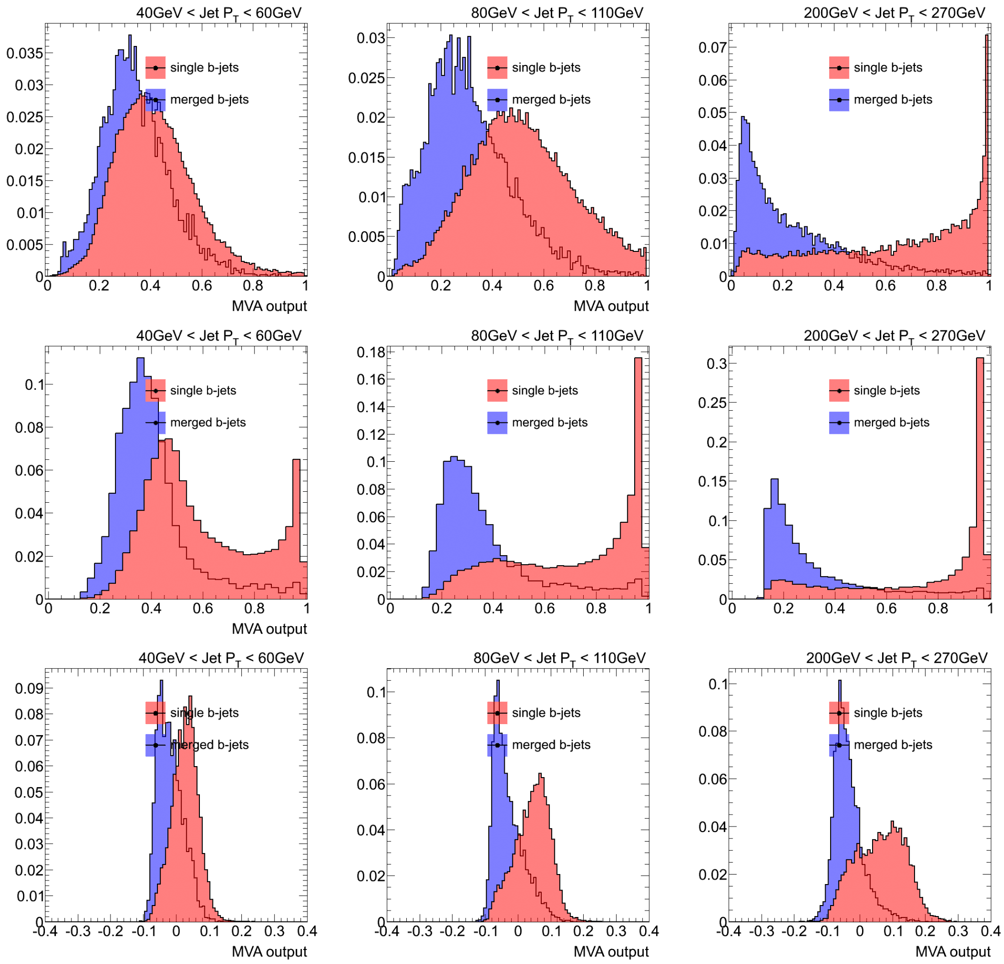

Figure 6.1: Distribution of the MVA discriminant outputs, for inclusive training, in single and merged b-jets, for low, medium and high jet p_T.

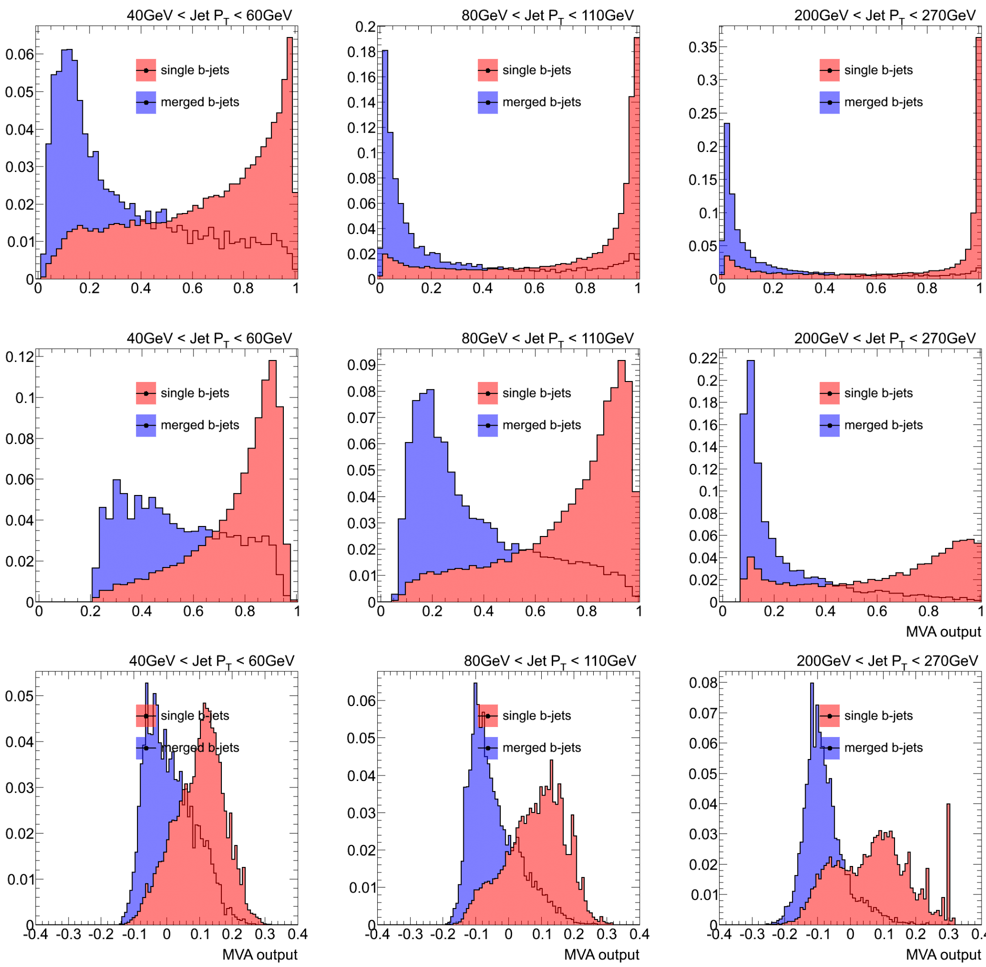

Figure 6.2: Distribution of the MVA discriminant outputs, for training in bins of jet p_T, in single and merged b-jets, for low, medium and high jet p_T.

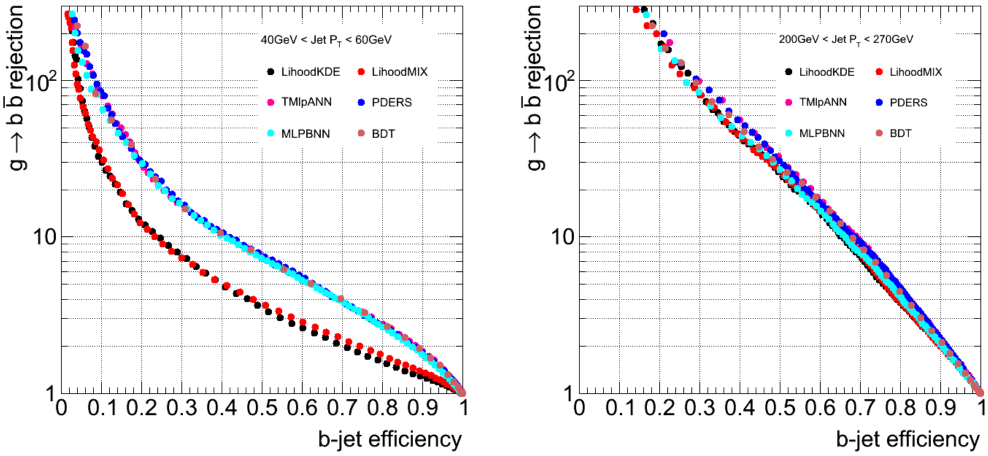

Figure 6.3: Distribution of the MVA discriminant performance for inclusive training, in single and merged b-jets, for low and high jet p_T.

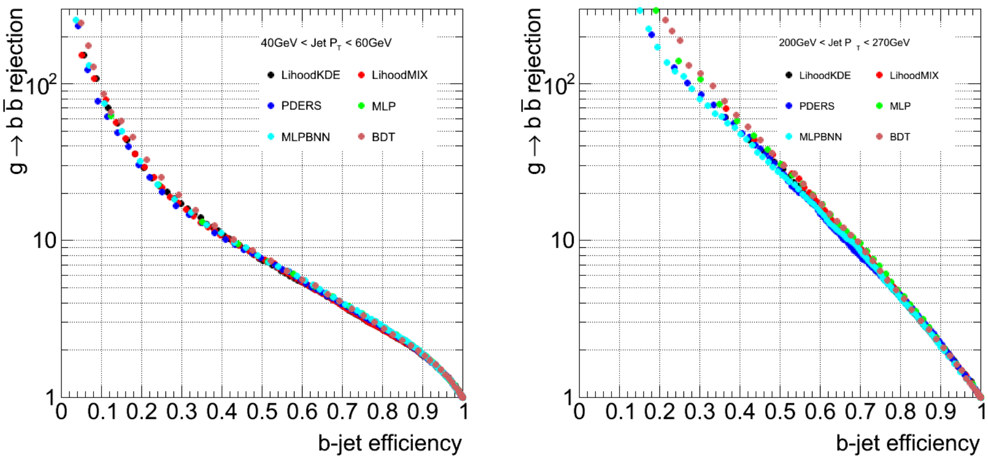

Figure 6.4: Distribution of the MVA discriminant performance for training in bins of jet p_T, in single and merged b-jets, for low and high jet p_T.

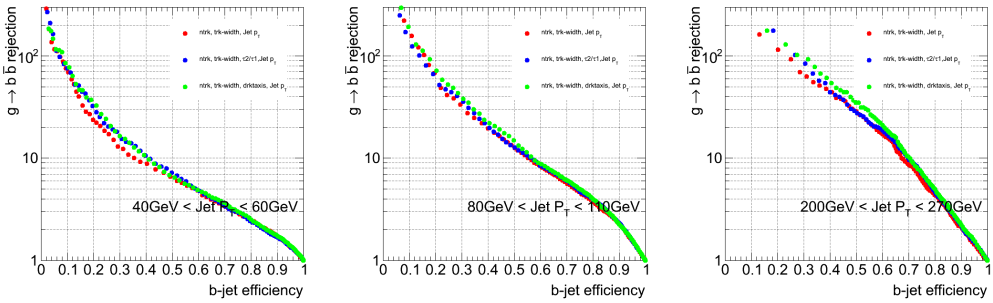

Figure 6.5: Distribution of the MVA discriminant performance for three sets of input variables, in single and merged b-jets, for low, medium and high jet p_T.

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 $g \to b\bar{b}$ 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 [82].

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.

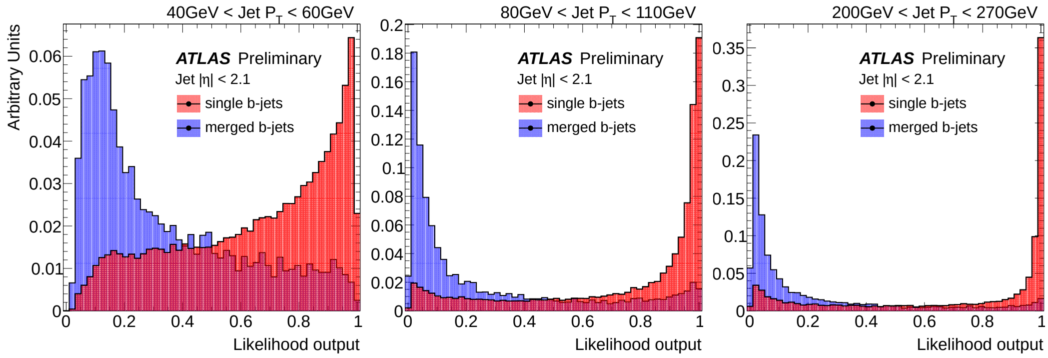

Figure 6.6: Distribution of the $g \to b\bar{b}$ likelihood output for single and merged b-jets for low, medium and high p_T jets.

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. ΔR 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 $g \to b\bar{b}$ 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 ϵ_{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.

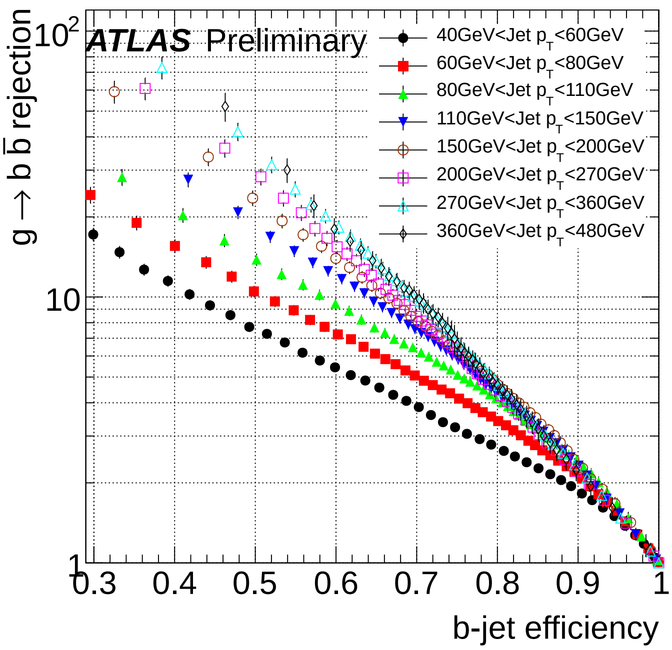

Figure 6.7: Rejection of $g \to b\bar{b}$ merged b-jets as a function of b-jet efficiency for dijet events in 8 jet p_T bins.

- $p_T > 60$ 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 $g \to b\bar{b}$ 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

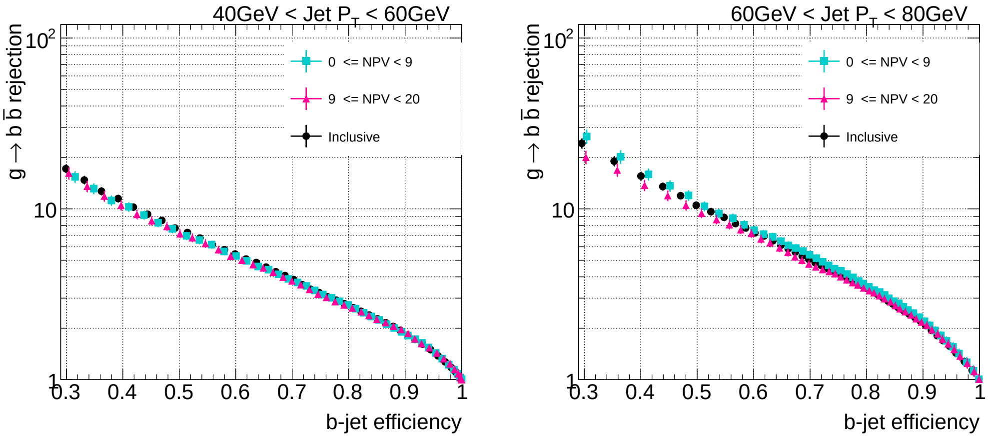

Figure 6.8: Rejection of $g \to b\bar{b}$ merged b-jets as a function of b-jet efficiency in bins of N_{vtx} for two low jet p_T bins.

- 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. [83].

The tracking efficiency systematics are given in bins of track η. For tracks with $p_{\mathrm{T}}^{\mathrm{track}} > 500$ MeV the uncertainties are independent of p_T: 2% for $|\eta^{\mathrm{track}}| < 1.3$, 3% for $1.3 < |\eta^{\mathrm{track}}| < 1.9$, 4% for $1.9 < |\eta^{\mathrm{track}}| < 2.1$, 4% for $2.1 < |\eta^{\mathrm{track}}| < 2.3$ and 7% for $2.3 < |\eta^{\mathrm{track}}| < 2.5$ [65]. 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

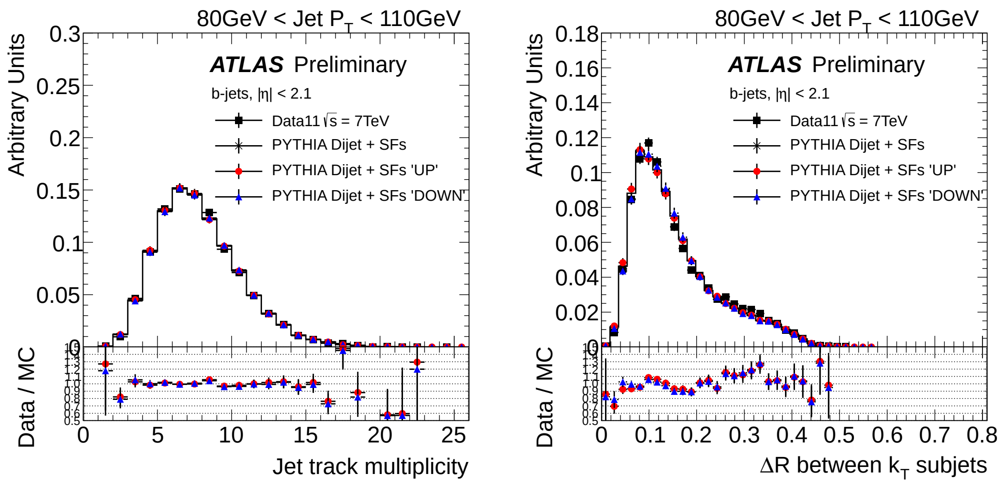

Figure 6.9: The effect of a variation in the b-tagging Scale Factors on the tracking variables distributions. Scale Factors were varied up (down) by 1-sigma to evaluate the systematic uncertainty from this source. The ratio data over MC is shown for MC PYTHIA with SFs varied up (circles) and down (triangles).

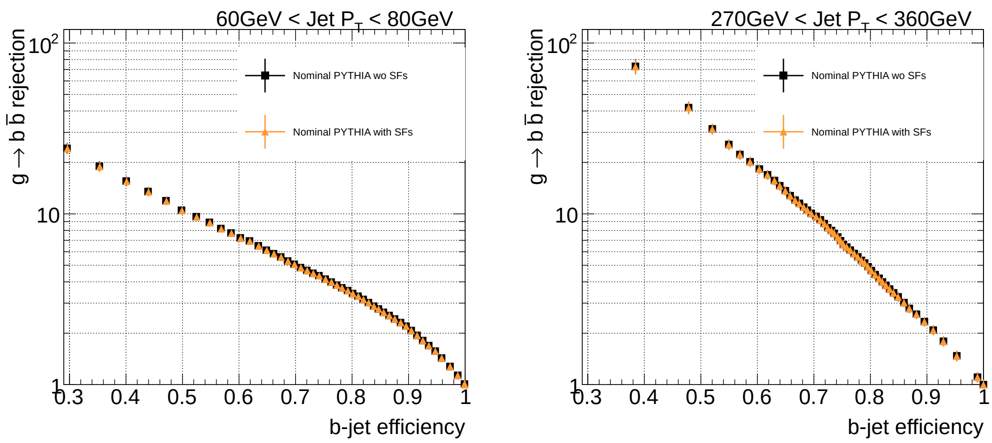

Figure 6.10: Rejection of $g \to b\bar{b}$ merged b-jets as a function of b-jet efficiency with and without scale factors.

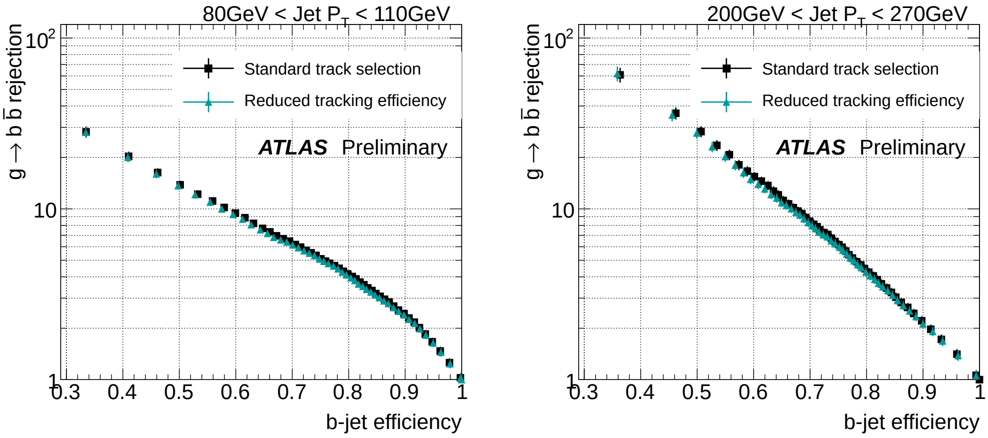

Figure 6.11: Rejection of $g \rightarrow b\bar{b}$ merged b-jets as a function of b-jet efficiency showing shift in likelihood performance caused by a reduction in the tracking efficiency .

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 over-smeared 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 [66]. 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 [84]. The precision of this measurement, determined in p_T and η 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 η or ϕ 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 $g \to b\bar{b}$ rejection are summarized in Table 6.2.

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

Table 6.2: Systematic uncertainities 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

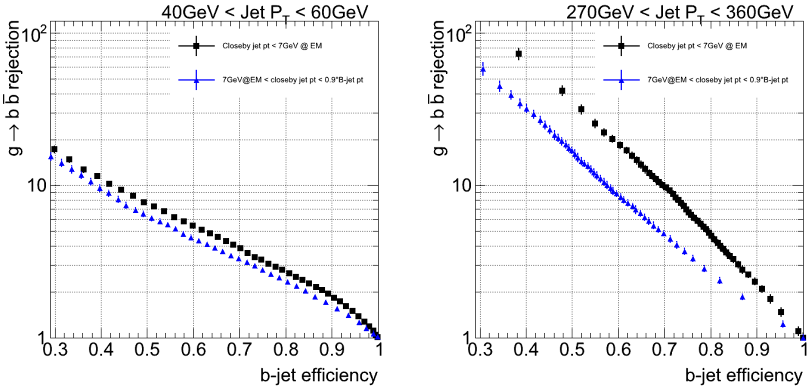

Figure 6.12: Rejection of $g \to b\bar{b}$ merged b-jets as a function of b-jet efficiency for two different isolation cuts.

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 susbtancial fraction of the b-jet pt 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.

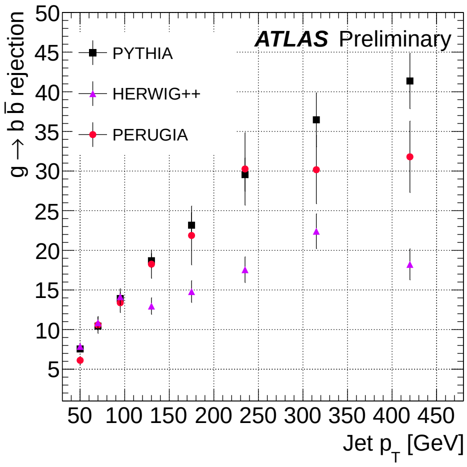

Figure 6.13: Rejection of $g \to b\bar{b}$ merged b-jets as a function of jet p_T for different Monte Carlo generators, at the 50% efficiency working point.

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 ΔR between the axes of the two k_t subjets in the jet, the two other variables used for discrimination.

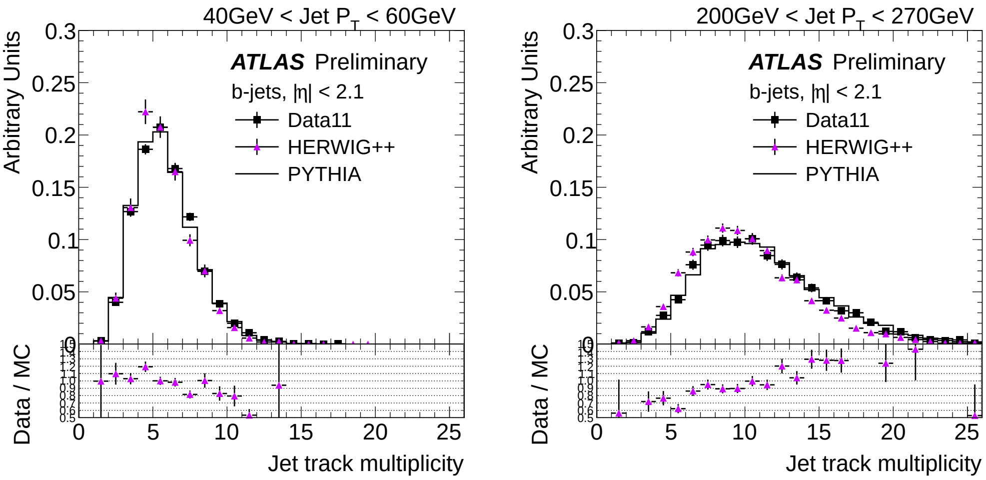

Figure 6.14: Distribution of the jet track multiplicity in 2 different jet p_T bins, for experimental data collected during 2011 (solid black points) and HERWIG++ events (solid violet triangules). The ratio data over HERWIG++ simulation is shown at the bottom of the plot. PYTHIA distribution is also 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 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 $p_T, achievingarejectionofmergedjetsofover 95\%(90\%) fora 50\% single\mathrm{b}-jete$ 150 GeV ($p_T > 60$ 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 discrimant variables, the extension to non-isolated jets using the concept of ghost-particle matching and active area of a jet [85] 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] S. Frixione and M.L. Mangano. Heavy quark jets in hadronic collisions. *Nucl.Phys.*, B483:321–338, 1997.

[2] Andrea Banfi, Gavin Salam, and Giulia Zanderighi. Accurate qcd predictions for heavy-quark jets at the tevatron and lhc. *JHEP*, 0707:026, 2007.

[3] Andrea Banfi, Gavin Salam, and Giulia Zanderighi. Infrared safe definition of jet flavour. *Eur.Phys.J.C*, 47:022, 2006.

[4] 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.

[5] 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.

[6] F. Abe et al. Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab.

[7] S. Abachi et al. Search for high mass top quark production in $p\overline{p}$ collisions at $\sqrt{s} =$ 1.8 tev.

[8] P.W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132 – 133, 1964.

[9] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. 2012.

[10] A. Sherstnev and R. S. Thorne. Parton distributions for LO generators. *European Physical Journal C*, 55:553–575, June 2008.

[11] F. J. Dyson. The radiation theories of tomonaga, schwinger, and feynman. *Phys. Rev.*, 75:486–502, Feb 1949.

[12] M Banner et al. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the cern pp collider. *Physics Letters B*, 122(56):476 – 485, 1983.

[13] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 gev/c2 at the cern sps collider. *Physics Letters B*, 126(5):398 – 410, 1983.

[14] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961.

[15] A. Salam and J.C. Ward. Electromagnetic and weak interactions. *Physics Letters*, 13(2):168 – 171, 1964.

[16] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[17] M. Gell-Mann. A schematic model of baryons and mesons. *Physics Letters*, 8(3):214 – 215, 1964.

[18] G. Zweig. AN SU(3) MODEL FOR STRONG INTERACTION SYMMETRY AND ITS BREAKING. 2. 1964.

[19] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. Advantages of the color octet gluon picture. *Physics Letters B*, 47(4):365 – 368, 1973.

[20] Richard P. Feynman. Very high-energy collisions of hadrons. *Phys. Rev. Lett.*, 23:1415–1417, Dec 1969.

[21] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv.Ser.Direct.High Energy Phys.*, 5:1–91, 1988.

[22] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298 – 318, 1977.

[23] G. Hanson, G. S. Abrams, A. M. Boyarski, M. Breidenbach, F. Bulos, W. Chinowsky, G. J. Feldman, C. E. Friedberg, D. Fryberger, G. Goldhaber, D. L. Hartill, B. Jean-Marie, J. A. Kadyk, R. R. Larsen, A. M. Litke, D. Lüke, B. A. Lulu, V. Lüth, H. L. Lynch, C. C. Morehouse, J. M. Paterson, M. L. Perl, F. M. Pierre, T. P. Pun, P. A. Rapidis, B. Richter, B. Sadoulet, R. F. Schwitters, W. Tanenbaum, G. H. Trilling, F. Vannucci, J. S. Whitaker, F. C. Winkelmann, and J. E. Wiss. Evidence for jet structure in hadron production by e^+e^- annihilation. *Phys. Rev. Lett.*, 35:1609–1612, Dec 1975.

[24] Michelangelo L. Mangano, Fulvio Piccinini, Antonio D. Polosa, Mauro Moretti, and Roberto Pittau. Alpgen, a generator for hard multiparton processes in hadronic collisions. *Journal of High Energy Physics*, 2003(07):001, 2003.

[25] Johan Alwall, Pavel Demin, Simon de Visscher, Rikkert Frederix, Michel Herquet, Fabio Maltoni, Tilman Plehn, David L. Rainwater, and Tim Stelzer. Madgraph/madevent v4: the new web generation. *Journal of High Energy Physics*, 2007(09):028, 2007.

[26] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[27] M Bhr, S. Gieseke, M.A. Gigg, A. Grellscheid, K. Hamilton, O. Latunde-Dada, S Pltzer, P Richardson, M.H Seymour, M Sherstnev, et al. Herwig++ physics and manual. *Eur.Phys.J.C*, 58:68, 2008.

[28] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(23):31 – 145, 1983.

[29] R. Corke and T. Sjöstrand. Improved parton showers at large transverse momenta. *European Physical Journal C*, 69:1, 2010.

[30] T. Sjöstrand and P. Z. Skands. Transverse-momentum-ordered showers and interleaved multiple interactions. *European Physical Journal C*, 39:129, 2005.

[31] Atlas tunes of pythia 6 and pythia 8 for mc11. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011.

[32] Peter Z. Skands. The Perugia Tunes. 2009.

[33] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev. D*, 82:074018, 2010.

[34] S. Agostinelli et al. Geant4 a simulation toolkit. *Nucl. Inst. Meth. Section A*, 506(3):250 – 303, 2003.

[35] Salam, G.P. Ellements of QCD for hadron colliders. *CERN-2010-002*, Jan 2011.

[36] Gavin P. Salam and Grgory Soyez. A practical seedless infrared-safe cone jet algorithm. *Journal of High Energy Physics*, 2007(05):086, 2007.

[37] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[38] W. Bartel, L. Becker, R. Felst, D. Haidt, G. Knies, H. Krehbiel, P. Laurikainen, N. Magnussen, R. Meinke, B. Naroska, et al. Experimental studies on multijet production in e^+e annihilation at PETRA energies. *EPJ C Particles and Fields*, 33:8, 1986.

[39] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.

[40] 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.

[41] M Cacciari and G.P. Salam. Dispelling the N^3 myth for the k_t jet-finder. *Phys. Lett. B}, 661:057, 2006.*

[42] G. Abbiendi et al. Measurement of alpha(s) with Radiative Hadronic Events. 2007.

[43] Georges Aad et al. Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at sqrt(s) = 7 TeV. 2012.

[44] *ATLAS Collaboration. Atlas sensitivity to the standard model higgs in the hw and hz channels at high transverse momenta.* ATL-PHYS-PUB-2009-088*, Aug 2009.*

[45] *Jason Gallicchio and Matthew D. Schwartz. Quark and gluon tagging at the lhc.* Phys. Rev. Lett.*, 107:172001, Oct 2011.*

[46] *Graham D. Kribs, Adam Martin, Tuhin S. Roy, and Michael Spannowsky. Discovering higgs bosons of the mssm using jet substructure.* Phys. Rev. D*, 82:095012, Nov 2010.*

[47] *Stephen Ellis, Christopher Vermilion, Jonathan Walsh, Andrew Hornig, and Christopher Lee. Jet shapes and jet algorithms in scet.* Journal of High Energy Physics*, 2010:1–83, 2010. 10.1007/JHEP11(2010)101.*

[48] *G. Aad et al. Study of jet shapes in inclusive jet production in pp collisions at* $\sqrt{s} = 7$ TeV *using the atlas detector.* Phys. Rev. D*, 83:052003, Mar 2011.*

[49] *Jason Gallicchio and Matthew D. Schwartz. Seeing in color: Jet superstructure.* Phys. Rev. Lett.*, 105:022001, Jul 2010.*

[50] *S. Catani, G. Turnock, and B.R. Webber. Jet broadening measures in e+e annihilation.* Physics Letters B*, 295(34):269 – 276, 1992.*

[51] *Leandro G. Almeida, Seung J. Lee, Gilad Perez, Ilmo Sung, and Joseph Virzi. Top quark jets at the lhc.* Phys. Rev. D*, 79:074012, Apr 2009.*

[52] *S. Catani, Yu.L. Dokshitzer, F. Fiorani, and B.R. Webber. Average number of jets in e+e annihilation.* Nuclear Physics B*, 377(3):445 – 460, 1992.*

[53] R. Snihur. Subjet multiplicity in quark and gluon jets at d0. Nuclear Physics B - Proceedings Supplements, 79(13):494 – 496, 1999. ¡ce:title¿Proceedings of the 7th International Workshop on Deep Inelastic Scattering and QCD¡/ce:title¿.

[54] A. Abdesselam, E. Bergeaas Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, et al. Boosted objects: A Probe of beyond the Standard Model physics. Eur.Phys.J., C71:1661, 2011.

[55] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Techniques for improved heavy particle searches with jet substructure. Phys. Rev. D, 80:051501, Sep 2009.

[56] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. JHEP, 1103:015:026, 2011.

[57] Iain W. Stewart, Frank J. Tackmann, and Wouter J. Waalewijn. n jettiness: An inclusive event shape to veto jets. Phys. Rev. Lett., 105:092002, Aug 2010.

[58] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. Phys. Rev. D, 48:3160–3166, Oct 1993.

[59] Jesse Thaler and Ken Van Tilburg. Maximizing boosted top identification by minimizing n-subjettiness. Journal of High Energy Physics, 2012:1–33, 2012. 10.1007/JHEP02(2012)093.

[60] Carola F. Berger, Tibor Kúcs, and George Sterman. Event shape ˘energy flow correlations. Phys. Rev. D, 68:014012, Jul 2003.

[61] Amos Breskin and Rdiger Voss. The CERN Large Hadron Collider: Accelerator and Experiments. CERN, Geneva, 2009.

[62] *ATLAS Collaboration. Luminosity Determination in pp Collisions at sqrt(s) = 7 TeV using the ATLAS Detector in 2011.* ATLAS-CONF-2011-116*, Aug 2011.*

[63] *S van der Meer. Calibration of the effective beam height in the isr.* CERN-ISR-PO-68-31. ISR-PO-68-31*, 1968.*

[64] *Aad, G. and others. The ATLAS Experiment at the CERN Large Hadron Collider. CERN, Geneva, 2008.*

[65] *G. Aad et al. Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC.* New J.Phys.*, 13:053033, 2011.*

[66] *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.*

[67] *ATLAS Collaboration. Tracking Studies for b-tagging with 7 TeV Collision Data with the ATLAS Detector.* ATLAS-CONF-2010-070*, 2010.*

[68] *W Lampl et al. Calorimeter Clustering Algorithms: Description and Performance. (ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003), Apr 2008.*

[69] *ATLAS Collaboration. Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data.* ATLAS-CONF-2012-020*, 2012.*

[70] *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.*

[71] G. Aad et al. *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics.* *2009.*

[72] ATLAS Collaboration. *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data.* ATLAS-CONF-2011-102, *2011.*

[73] ATLAS Collaboration. *Calibrating the b-Tag Efficiency and Mistag Rate in 35 pb^{-1} of Data with the ATLAS Detector.* ATLAS-CONF-2011-089, *2011.*

[74] 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.*

[75] ATLAS Collaboration. *Performance of the ATLAS Secondary Vertex b-tagging Algorithm in 7 TeV Collision Data.* ATLAS-CONF-2010-042, *2010.*

[76] ATLAS Collaboration.

[77] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. *A brief introduction to pythia 8.1.* Comput. Phys. Commun., *178:852, 2008.*

[78] B. Andersson et al. *Parton fragmentation and string dynamics.* Phys. Rep., *97:31, 1983.*

[79] ATLAS Collaboration. *The ATLAS Simulation Infrastructure.* Eur.Phys.J.C, *70:051, 2010.*

[80] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. *HERWIG 6: An Event generator for hadron emission reactions with in-*

terfering gluons (including supersymmetric processes). JHEP*, 0101:010, 2001.*

[81] Manuel Bahr, Stefan Gieseke, and Michael H. Seymour. Simulation of multiple partonic interactions in Herwig++. JHEP*, 0807:076, 2008.*

[82] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. PoS*, ACAT:040, 2007.*

[83] 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.*

[84] G. Romeo, A. Schwartzman, R. Piegaia, 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.*

[85] G.P. Salam M. Cacciari and Gregory Soyez. The Catchment Area of Jets. JHEP*, 0804:42, 2008.*