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

1 Event reconstruction and b-Tagging **2**

1.1 Jet reconstruction and calibration 3

1.2 Reconstruction of charged particle tracks 13

1.3 Vertex reconstruction 15

1.4 b-jet Tagging 16

1.4.1 b-tagging algorithms 19

Chapter 1

Event reconstruction and b-Tagging

The event reconstruction software, which in ATLAS is implemented in the software framework ATHENA, process the events starting from the raw data obtained from the various sub-detectors (energy deposits and hits), processing them in many different stages and finally interpreting them as a set of charged tracks, electrons, photons, jets, muons and, in general, of possible kinds of final state objects with related four momenta. In this chapter the reconstruction of these objects is briefly described together with the algorithms for the identification of b-quark jets. These algorithms are mainly based on the reconstruction of the primary interaction vertex, on the reconstruction of charged particles in the Inner Detector and on the reconstruction of jets in the calorimeter.

1.1 Jet reconstruction and calibration

Hadronic jets used for ATLAS analyses are reconstructed by a jet algorithm, starting from the energy depositions of electromagnetic and hadronic showers in the calorimeters. Two different size parameters are used: $R = 0.4$, for narrow jets, and $R = 0.6$, for wider jets. The default jet algorithm is the anti-kt algorithm, described in the previous chapter. Due to the expected level of pile-up in the LHC, the primary factor that influenced the selection of this algorithm was the effect of multiple simultaneous interactions on the reconstruction of jets. The original ATLAS cone algorithm, known to contain infrared and collinear sensitivity, is highly susceptible to this effect. On the contrary, the anti-kt algorithm is the most stable after the introduction of pile-up [1].

The input to calorimeter jet reconstruction can be calorimeter towers or topological cell clusters. Charged particle tracks reconstructed in the Inner Detectors are also used to define jets. These constituents have the further advantage of being insensitive to pile-up and they provide a stable reference for systematic studies. The jet inputs are combined as massless four-momentum objects in order to form the final four-momentum of the jet, which allows for a well-defined jet mass [2]. In the case of track-jets, the track four-momentum is constructed assuming the π meson mass for each track.

Calorimeter towers are static, $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$, grid elements built directly from calorimeter cells. There are two types of calorimeter towers: with or without noise supression. The latter are called "noise-suppressed" towers and use only the cells with energies above a certain noise threshold. The noise of a calorimeter cell is measured by recording calorimeter signals in periods where no beam is present in the acelerator. The standard deviation

σ around the mean measured energy is interpreted as the noise of the cell, and dependes on the sampling layer in which the cell resides and the position in η.

The results presented in this thesis show jets which were built from noise-suppressed topological clusters of energy in the calorimeter, aslo known as "topo-clusters" [3]. Topological clusters are groups of calorimeter cells that are designed to follow the shower development taking advantage of the fine segmentation of the ATLAS calorimeters. The topological cluster formation starts from a seed cell with $|E_{cell}| > 4\sigma$ above the noise. In a second step, neighbor cells that have an energy at least 2σ above their mean noise are added to the cluster. Finally, all nearest-neighbor cells surrounding the clustered cells are added to the cluster, regardless of signal-to-noise ratio[1]. The position of the cluster is assigned as the energy-weighted centroid of all constituent cells (the weight used is the absolute cell energy).

In Monte Carlo simulation, reference jets ("truth jets") are formed from simulated stable particles using the jet algorithm utilized for the reconstructed jets.

Jet calibration

The purpose of the jet energy calibration, or jet energy scale (JES), is to correct the measured electromagnetic scale (EM scale) energy to the energy of the stable particles within a jet. The jet energy calibration must account then for the calorimeter non-compensation; the energy lost in inactive regions of the detector, such as in the cryostat walls or cabling; energy that escapes the calorimeters, such as that of highly-energetic particles that "punch-through"

[1]Noise-supressed towers also make use of the topological clusters algorithm [3] to select cells, i.e. only calorimeter cells that are included in topo-clusters are used.

to the muon system; energy of cells that are not included in clusters, due to inefficiencies in the noise-suppression scheme; and energy of clusters not included in the final reconstructed jet, due to inefficiencies in the jet reconstruction algorithm. The muons and neutrinos that may be present within the jet are not expected to interact within the calorimeters, and are not included in this energy calibration. Due to the varying calorimeter coverage, detector technology, and amount of upstream inactive material, the calibration that must be applied to each jet to bring it to the hadronic scale varies with its η position within the detector.

The jet energy is first reconstructed from the constituent cell energies at EM scale. These cells have been calibrated to return the energy corresponding to electromagnetic showers in the calorimeter, based on test-beam injection of electrons and pions [4], measurements of cosmic muons [5] and the reconstruction of the Z mass peak in $Z \to ee$ decays [6]. The correction for the lower response to hadrons is based on the topology of the energy depositions observed in the calorimeter.

In the simplest case the measured jet energy is corrected, on average, using Monte Carlo simulations, as follows:

$$E^{jet}_{calib} = E^{jet}_{meas}/F_{calib}(E^{jet}_{meas}), \text{with } E^{jet}_{meas} = E^{jet}_{EM} - O(\text{NPV}), \tag{1.1}$$

where E^{jet}_{EM} is the calorimeter energy measured at the electromagnetic scale, E^{jet}_{calib} is the calibrated energy and F_{calib} is the calibration function that depends on the measured jet energy and is evaluated in small jet η regions. The variable $O(NPV)$ denotes the correction for additional energy from multiple proton-proton interactions depending on the number of primary vertices (NPV).

The simplest calibration scheme and the one used in this thesis is the so called "EM+JES". This calibration applies the corrections as a function of the

jet energy and pseudorapidity to jets reconstructed at the electromagnetic scale. The additional energy due to multiple proton-proton collisions within the same bunch crossing (pile-up) is corrected before the hadronic energy scale is restored, such that the derivation of the jet energy scale calibration is factorised and does not depend on the number of additional interactions in the event. The EM+JES calibration scheme consists of three subsequent stops:

- Pile-up correction: An offset correction is applied in order to substract the additional average energy measured in the calorimeter due to multiple proton-proton interactions. This correction is derived from minimum bias data as a function of NPV, the jet pseudorapidity and the bunch spacing.

- Vertex correction: The jet four momentum is corrected such that the jet originates from the primary vertex of the interaction instead of the geometrical centre of the detector.

- Jet energy and direction correction: The jet energy and direction are corrected using constants derived from the comparison of the kinematic observables of reconstructed jets and those from truth jets in the simulation.

In the final step the calibration is derived in terms of the energy response of the jet, or the ratio of the reconstructed jet energy to that of a truth jet. The EM scale response is written as,

$$R_{EM}^{jet} = E_{EM}^{jet} / E_{truth}^{jet} \tag{1.2}$$

To compute this quantity, reconstructed jets must be matched to isolated jets in the Monte Carlo within $\Delta R < 0.3$. The isolation requirement is

applied in order to factorize the effects due to close-by jets from those due to purely detector effects such as dead material and non-compensation. The isolation criterion requires that no other jet with a $p_T > 7$ GeV be within $\Delta R < 2.5R$, where R is the distance parameter of the jet algorithm. The EM scale energy response is binned in truth jet energy, E^{jet}_{truth} and the calorimeter jet detector η. For each (E^{jet}_{truth}, η)-bin, the averaged jet response is defined as the peak position of a Gaussian fit to the $E^{jet}_{EM}/E^{jet}_{truth}$ distribution. A function $F_{calib,k}(E^{jet}_{EM})$ is then defined for each η-bin k that describes the response as a function of the uncalibrated jet energy. $F_{calib,k}(E^{jet}_{EM})$ is parameterised as:

$$F_{calib,k}(E^{jet}_{EM}) = \sum_{i=0}^{N_{max}} a_i (\ln E^{jet}_{EM})^i, \tag{1.3}$$

where a_i are free parameters, and N_{max} is chosen between 1 and 6 depending on the goodness of the fit. The final jet energy scale correction that relates the measured calorimeter jet energy scale to the hadronic scale is then defined as $1/F_{calib,k}(E^{jet}_{EM})$ in the following:

$$E^{jet}_{EM+JES} = \frac{E^{jet}_{EM}}{F_{calib}(E^{jet}_{EM})|_\eta}, \tag{1.4}$$

where $F_{calib}(E^{jet}_{EM})|_\eta$ is $F_{calib,k}(E^{jet}_{EM})$ for the relevant η-bin k.

Other calibrations schemes are the global calorimeter cell weighting (GCW) calibration and the local cluster weighting (LCW) calibration. The GCW scheme exploits the observation that electromagnetic showers in the calorimeter leave more compact energy depositions than hadronic showers with the same energy. Energy corrections are derived for each cell within a jet. The cell corrections account for all energy loses of a jet in the detector. Since these corrections are only applicable to jets and not to energy depositions, they are called "global" corrections.

The LCW calibration method first classifies topo-clusters as either electromagnetic of hadronic, based on the measured energy density. Energy

corrections are derived according to this classification from single charged and neutral pion Monte Carlo simulations. Dedicated corrections are derived for the effects of non-compensation, signal losses due to noise threshold effects, and energy lost in non-instrumented regions. Since the energy corrections are applied without reference to a jet definition they are called "local" corrections. Jets are then built from these calibrated clusters using a jet algorithm.

The final jet energy calibration can be applied to EM scale jets, with the resulting calibrated jets referred to as EM+JES, or to LCW (GCW) calibrated jets, with the resulting jets referred to as LCW+JES (GCW+JES) jets.

A further jet calibration scheme called global sequential (GS) calibration, starts from jets calibrated with the EM+JES calibration and exploits the topology of the energy deposits in the calorimeter to characterise fluctuations in the jet particle content of the hadronic shower development. Correcting for such fluctuations can improve the jet energy resolution. The correction uses several jet properties, and each correction is applied sequentially.

For the 2011 data the recomended calibration schemes were the EM+JES and the LCW calibrations. The simple EM+JES calibration does not provide the best performance, but allows in the central detector region the most direct evaluation of the systematic uncertainties from the calorimeter response to single isolated hadrons measured *in situ* and in test-beams and from systematic variations in the Monte Carlo simulation. For the LCW+JES calibration scheme the JES uncertainty is determined from *in situ* techniques. For all calibration schemes, the JES uncertainty in the forward regions is derived from the uncertainty in the central region using the transverse momentum balance in events where only two jets are produced.

Jet energy scale uncertainties for the EM+JES scheme

For many physics analyses, the uncertainty on the JES constitutes the dominant systematic uncertainty because of its tendency to shift jets in and out of analysis selections due to the steeply falling jet p_T spectrum. The uncertainty on the EM+JES scale is determined primarily by six factors: varying the physics models for hadronization and parameters of the Monte Carlo generators, evaluating the baseline calorimeter response to single particles, comparing multiple models for the detector simulation of hadronic showers, assesing the calibration scales as a function of pseudorapidity, and by adjusting the JES calibration methods itself. The final JES uncertainty in the central region, $|\eta| < 0.8$, is determined from the maximum deviation in response observed with respect to the response in the nominal sample. For the more forward region, the so called "η-intercalibration" contribution is estimated. This is a procedure that uses direct di-jet balance measurements in two-jet events to measure the relative energy scale of jets in the more forward regions compared to jets in a reference region. The technique exploits the fact that these jets are expected to have equal p_T due to transverse momentum conservation. Figure 1.1 shows the final fractional jet energy scale uncertainty and its individual contributions as a function of p_T for a central η region. The JES uncertainty for anti-kt jets with $R = 0.4$ is between $\approx$4% (8%, 14%) at low jet p_T and $\approx$2.5%-3% (2.5%-3.5%, 5%) for jets with $p_T > 60$ GeV in the central (endcap, forward) region.

In addition to the tests above, *in situ* tests of the JES using direct γ-jet balance, multi-jet balance, and track-jets indicate that the uncertainties in Fig. 1.1 reflect accurately the true uncertainties in the JES.

In the case of jets induced by bottom quarks (b-jets), the calorimeter response uncertainties are also evaluated using single hadron response mea-

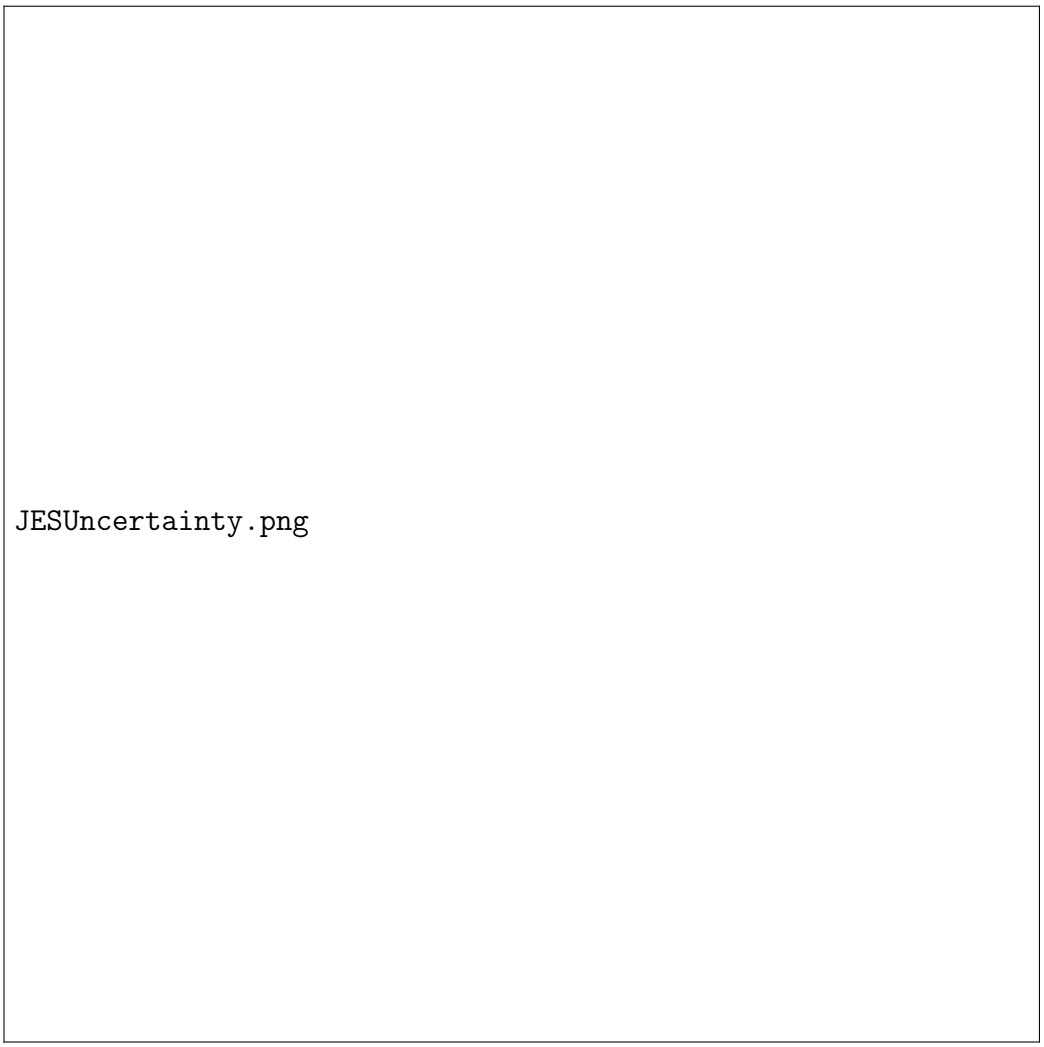

Figure 1.1: Fractional jet energy scale uncertainty as a function of jet p_T for jets in the pseudorapidity region $0.3< |\eta| <0.8$ in the calorimeter barrel. The total uncertainty is shown as the solid tight blue area. The individual sources are also shown.

surements *in situ* and in test beams [7]. For jets within $|\eta| < 0.8$ and $20 \leq p_T < 250$ GeV the expected difference in the calorimeter response uncertainty of identified b-jets with respect to the one of inclusive jets is less than 0.5%. It is assumed that this uncertainty extends up to $\eta < 2.5$.

The JES uncertainty arising from the modelling of the b-quark fragmentation can be determined from systematics variations of the Monte Carlo simulation. The fragmentation function is used to estimate the momentum carried by the B-hadron with respect to that of the b-quark after quark fragmentation. The fragmentation function included in PYTHIA originates from a detailed study of the b-quark fragmentation function in comparison with OPAL [8] and SLD [9] data. To assess the impact of the b-fragmentation, the nominal parameters of the PYTHIA fragmentation function are replaced by the values from a tune using the Professor framework [10]. In addition, the nominal fragmentation function is replaced by the modified Bowler-Lund fragmentation function [11]. The b-jet response uncertainty is evaluated from the ratio between the response of b-jets in the varied Monte Carlo samples to the nominal PYTHIA. The response variations are well within 2%.

The b-jet JES uncertainty is obtained adding the calorimeter response uncertainty and the uncertainties from the systematic Monte Carlo variations in quadrature. The resulting additional JES uncertainty for b-jets is shown in Fig. 1.2. It is about 2% up to $p_T \approx 100$ GeV and below 1% for higher p_T. To obtain the overall b-jet uncertainty this uncertainty is added in quadrature to the JES uncertainty for inclusive jets.

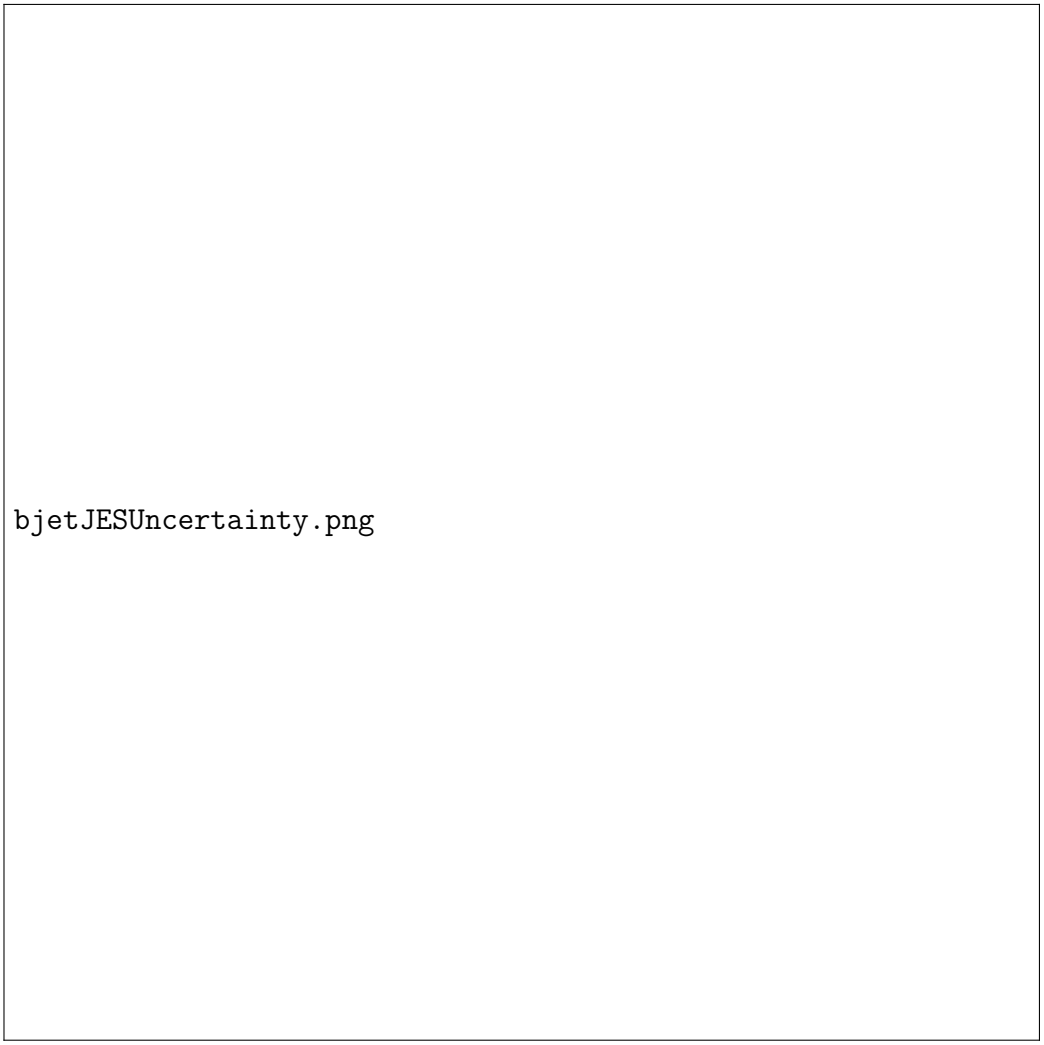

Figure 1.2: Additional fractional b-jet JES uncertainty as a function of the truth jet transverse momentum for anti-kt jets with $R = 0.4$ calibrated with EM+JES scheme for $|\eta| < 2.5$. Shown are systematic Monte Carlo variations using different modelling of the b-quark fragmentation and physics effects as well as variations in the detector geometry and the uncertainty in the calorimeter response to b-jets as evaluated from single hadron response measurements. Uncertainties in the individual points are statistical only.

1.2 Reconstruction of charged particle tracks

The Inner Detector layout and the characteristics of its main sub-detectors were presented in Section **??** of Chapter **??**. The tracking algorithm is based on a modular software framework, which is described in more detail in Ref. [12]. The main steps of the tracking algorithm are the following:

- Firstly, the raw data from the pixel and SCT detectors are converted into clusters, while the TRT raw timing information is turned into calibrated drift circles. The SCT clusters need to be further transformed into space-points, by combining the clusters information from opposite sides of the SCT module (stereo strip layers).

- In a second stage, the track-finding is performed, in which the pattern recognition and a global χ^2 minimization procedure is implemented as a default.

In the track-finding stage, track seeds are found in the first three pixel layers and in the first SCT layer. These are extended throughout the SCT to form track candidates and a first track fit is performed. Afterwards, ambiguities in the track candidates found in the silicon detectors are resolved, and tracks are extended into the TRT (which covers up to $|\eta| < 2.$). The final track candidate is refitted with the full information from the three tracking subdetectors. The baseline algorithm is designed for the efficient reconstruction of primary charged particles. Primary particles are defined as particles with a meanlife of greater than 3×10^{-11} s directly produced in a proton-proton interaction, or from the subsequent decays or interactions of particles with lifetime shorter than 3×10^{-11} s. The tracks reconstructed in this stage are required to have $p_T > 400$ MeV.

In a complementary stage, a track search starts from segments reconstructed in the TRT and extends them inwards by adding silicon hits, which is referred to as "back-tracking". This recovers tracks for which the first hits in the pixel layers are missing, e.g. because they originate from secondaries, which are produced in the interaction of primaries.

The final reconstructed track trajectory is usually defined at its closest point to the interaction region on the transverse plane by its impact parameters in the transverse plane and in the longitudinal direction, respectively called d_0 and ${z_0}^2$., and by its momentum, typically expressed in azimuthal angle ϕ, polar angle θ and inverse momentum $1/p$.

The track reconstruction efficiency is defined as the fraction of primary particles with $p_T > 400$ MeV and $|\eta| < 2.5$ matched to a reconstructed track. The reconstruction efficiency for primary tracks with transverse momentum above 1 GeV and central η is above 80%, reaching values below 70% for tracks at the edge of the Inner Detector acceptance. The dense environment of a jet decreases the track reconstruction efficiency and increases the fake rate. This is caused by the ocurrence of shared hits between different tracks, which makes the pattern recognition and track fitting tasks more difficult.

The relative transverse momentum scale and resolution of tracks is defined as the Gaussian mean and width of

$$ {p_T}^{MC} \times (1/{p_T}^{MC} - 1/{p_T}^{reco}) = 1 - \frac{{p_T}^{MC}}{{p_T}^{reco}} \tag{1.5} $$

where ${p_T}^{MC,reco}$, refers to the track's transverse momentum given by simulation truth (MC) or by reconstruction (reco). It should be noted that the $(1/p_T)$ resolution is used instead of $\sigma(p_T)$ as the Inner Detector measures the

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

sagitta and not directly the transverse momentum[3]. However, the resolution obtained from the equation above is the relative transverse momentum resolution, $\sigma(p_T)/p_T$. At low p_T the multiple scattering dominates the resolution, and at high momenta, the resolution is limited by the bending power of the solenoid field and by the instrinsic detector resolution. For a central track with p_T =5 GeV the transverse momentum resolution is around 75 MeV and the transverse impact parameter resolution is about 35 μm.

1.3 Vertex reconstruction

Primary vertices are reconstructed using an iterative vertex finding algorithm [13]. In a first step, a dedicated vertex finding algorithm associates tracks to vertex candidates. Vertex seeds are obtained by looking for the global maximum in the distribution of the z coordinates of the tracks. In a second stage, an iterative χ^2 fit is made using the seed and nearby tracks. Each track carries a weight which is a measure of its compatibility with the fitted vertex depending on the χ^2 of the fit. Tracks displaced by more than 7σ from the vertex are used to seed a new vertex and the procedure is repeated until no additional verteces can be found. The parameters of the beam spot are used both during the finding to preselect compatible tracks and during the fitting step to constrain the vertex fit.

The knowledge of the position of the primary interaction point (primary vertex) of the proton-proton collision is important for b-quark jets identification since it defines the reference point with respect to which impact parameters and vertex displacements are measured. The typical vertexing resolution in z is $O(100\mu\text{m})$.

[3]The relation between sagitta s and transverse momentum (p_T) is given by $s \sim 1/p_T$.

To ensure a good resolution on the vertex position, the primary vertex must be reconstructed from at least five tracks. The choice of the primary vertex is less trivial in the presence of minimum-bias events from pile-up: the primary vertex from a pile-up event may be mistakenly used as the signal vertex, or a fake primary vertex built from tracks from two different vertices may be reconstructed. The current strategy is to choose the primary vertex candidate that maximizes $\sum_{tracks} {p_T}^2$.

1.4 *b*-jet Tagging

The ability to identify jets originating from *bottom*-quarks (denoted as *b*-tagging in the following) is important for the high-p_T physics program of a general-purpose experiment at the LHC such as ATLAS since many interesting physics processes contain *b*-quarks in the final, state while the most abundant backgrounds contain mostly up, down and strange quark or gluon jets or, in a smaller fraction of cases, charm quark jets. The aim of *b*-tagging is therefore to identify the *b*-quark jets with high efficiency, while rejecting most of the background contamination from jets originating from the fragmentation of light (u, d, and s) quarks, gluons and *c*-quarks.

A *b*-quark, once produced, fragments necessarily into a *b*-flavoured hadron, *b*-hadron in the following. In most of the cases ($\approx$87%), first an excited *b*-hadron, like a B^* or a B^{**}, which decays inmediately, strongly or electromagnetically, into a ground state *b*-hadron plus one or more further particles; while in the remaining cases, a ground state *b*-hadron is produced directly. One is only interested in the transition from a *b*-quark into the final state *b*-hadron, since the typical timescale for electromagnetic or strong interactions is so small that the B^*, B^{**} decay vertices are not significantly displaced

with respect to the primary vertex. In most of the cases ($\approx 91\%$) a B-meson is produced out of the fragmentation of an original b-quark (40% B^+, 40% B^0 and 11% B_s^0). The rest are b-baryons.

Due to the b-quark fragmentation function being very hard, most of the original b-quark energy is transmitted to the final b-hadron. This fraction is for example 70% for b-quarks with a momentum of $\approx$45 GeV. This property can be exploited during b-tagging, since the fragmentation for light quarks into light hadrons or c-quarks into c-hadrons is softer.

Any of the finally produced b-hadrons decay through weak interactions and therefore have a significant lifetime, which is on average, for all b-hadrons considered, $(1.568 \pm 0.009) \times 10^{-12}$ s. The effective distance travelled in the detector by the b-hadron before decaying depends on the b-hadron momentum, which enters the relativistic boost factor $\beta\gamma$. A b-quark with momentum of 50 GeV will travel around 3 mm, which is a visible flight length in the detector. Due to the combination of the b-hadron lifetime and relatively high mass ($m_B \approx 5.28$ GeV), which results in a non-negligible decay angle of the b-hadron decay products with respect to the b-hadron flight direction, the charged particles produced at the decay vertex will be on average significantly displaced with respect to the primary vertex position.

This is the main signature which is exploited by the $lifetime$ based b-tagging algorithms, which are based either on the presence of significantly displaced tracks, as in impact parameter based b-tagging algorithms, or on the explicit reconstruction of the b-hadron decay vertex, as in secondary vertex based b-tagging algorithms.

b-hadrons decay preferably into a c-hadron plus additional particles[4]. The lifetime of a c-hadron is not much lower than for b-hadrons, but in general

[4]Weak decays are governed by the CKM matrix mechanism, and $|V_{cb}|^2 >> |V_{ub}|^2$.

the momentum of the c-hadron will be lower than the original b-hadron momentum. However, the c-hadron can still travel for a significant path in the detector and form with its decay producs a visible *tertiary* vertex.

Another property which is usually exploited by b-tagging is the fraction of b- and c-hadron decays into leptons: a lepton from the semi-leptonic decay of a b-hadron ($b \to l$) or from the subsequent c-hadron decay ($b \to c \to l$) is produced in a b-quark in $\approx$21% of the cases. This is valid both in case the lepton is a an electron or a muon, which brings the overall fraction of b-quarks ending up into a lepton to $\approx$42%. Due to the b- or c-hadron mass, the lepton will be emitted with an average transverse momentum comparable with m_{b-had} or m_{c-had}. By identifying either an electron or a muon originating from a jet and by requiring it to have sufficiently high p_T with respect to the jet axis, it is possible to identify b-jets.

Association of tracks to jets

The b-tagging performance relies criticaly on the accurate reconstruction of the charged tracks in the ATLAS Inner Detector. The actual tagging is performed on the sub-set of tracks in the event that are associated to jets. The b-tagging algorithm takes as input the three-momenta of the jets, reconstructed by a jet algorithm, and uses the jet direction to associate the charged particles tracks to the jet. Since the 2 Tesla solenoidal magnetic field of the ATLAS Inner Detector bends charged particles in the transverse plane, in particular in the case of low p_T tracks, the tracks are best matched to the jet by using the direction of their momenta at the point of closest approach to the interaction region. The criterion for associating charged particle tracks to jets is simply:

$$\Delta R(jet, track) < \Delta R_{cut} \tag{1.6}$$

where usually the value of $\Delta R_{cut} = R$ is used; with R, the distance parameter of the jet algorithm used for jet reconstruction.

After the tracks are associated to the jets, they are filtered in order to remove tracks with bad quality or which can easily be erroneously identified as secondary tracks from b-decays. These include tracks originating from decays of even longer lived particles, like K_s^0 ($c\tau \approx$2.69 cm) and Λ baryons ($c\tau \approx$7.89 cm); from electromagnetic interactions in the detector material, like conversions in electron-positron pairs ($\gamma \to e^+e^-$); or from hadronic interactions with the detector material, which result in two or more tracks with high impact parameter. In order to reject badly reconstructed tracks, quality cuts are applied. Requirements are imposed on the number of silicon hits, the track fit quality, the track momentum, and the transverse and longitudinal impact parameters. The track selection needs to be particularly tight in the case of the impact parameter based b-tagging algorithms, since in that case the explicit presence of a vertex is not required, so that the influence of badly reconstructed tracks or tracks from long lived particles does directly limit the performance. The minimum track p_T required is of 1 GeV in the case of the impact parameter based algorithms and of 400-500 MeV otherwise. The transverse and longitudinal impact parameters must fullfill $|d_0|$ <1 mm (3.5 mm) and $|z_0| \sin\theta < 1.5$ mm (no cut on z_0) in the case of the algorithms relying on the impact parameters of tracks (on the reconstruction of secondary vertices). The minimum number of precision hits required is typically of 7 hits, for both approaches.

1.4.1 b-tagging algorithms

For the 2011 data-taking a set of lifetime taggers were commissioned and calibrated. In this section a brief description of the main features of these

algorithms will be given.

Impact parameter based b-tagging algorithms

The charged particle tracks originating from b-hadrons are expected to have significantly higher transverse and longitudinal impact parameters compared to prompt tracks originating directly from fragmentation. If the effect of long lived particles, conversions and hadronic interactions can be reduced, the best discrimination between prompt tracks and displaced tracks from b- and c-hadron decays can be obtained using the impact parameter significance, both in the transverse and longitudianl plane. Being,

$$IP_{r\phi} = d_0 \;\; and \;\; IP_z = z_0 \sin\theta, \tag{1.7}$$

The transverse and longitudinal impact parameter significances are obtained by dividing $IP_{r\phi}$ and IP_z by their respective errors,

$$IP_{r\phi}/\sigma(IP_{r\phi}) \;\; and \;\; IP_z/\sigma(IP_z). \tag{1.8}$$

On the basis that the decay point of the b-hadron must lie along its flight path, and in order to increase the discriminating power of the impact parameter significance, a lifetime sign is assigned to these variables (replacing the sign of the geometrical definition of the impact parameter). The sign is positive if the track extrapolation crosses the jet direction in front of the primary vertex (i.e. is more compatible with having its origin in a secondary decay vertex in the direction of flight expected for the b-hadron) or negative if the track is more likely to intersect the flight axis behind the primary vertex, opposite to the jet direction. Both cases are illustrated in Fig. 1.3.

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

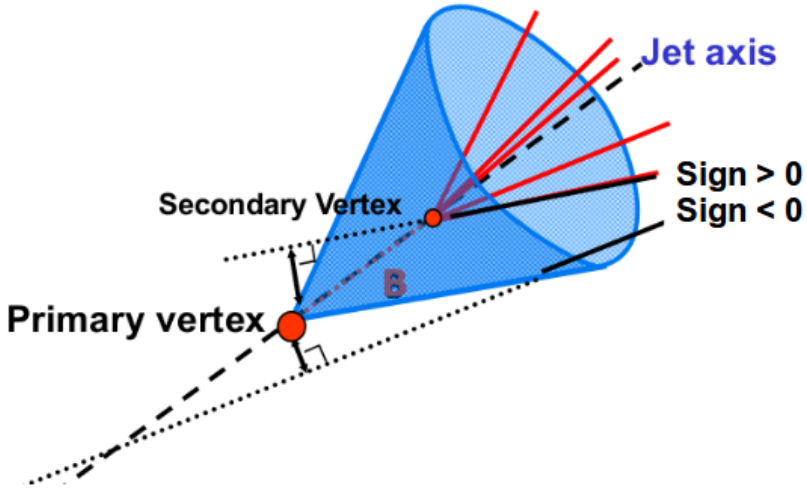

Figure 1.3: Lifetime sign of tracks. A positive and a negative lifetime signed track is shown.

Bibliography

[1] ATLAS Collaboration.

[2] ATLAS Collaboration.

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

[4] M. Aharrouche et al. Energy linearity and resolution of the atlas electromagnetic barrel calorimeter in an electron test-beam. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 568(2):601 – 623, 2006.

[5] M Cooke, P S Mangeard, M Plamondon, M Aleksa, M Delmastro, L Fayard, S Henrot-VersillÃľ, F Hubaut, R Lafaye, W Lampl, J LÃľvÃłque, H Ma, E Monnier, J Parsons, P Pralavorio, Ph Schwemling, L Serin, B TrocmÃľ, G Unal, M Vincter, and H Wilkens. In situ commissioning of the ATLAS electromagnetic calorimeter with cosmic muons. *ATL-LARG-PUB-2007-013*, 2007.

[6] Georges Aad et al. Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data. *Eur.Phys.J.*, C72:1909, 2012.

[7] ATLAS Collaboration.

[8] G. Abbiendi et al. Inclusive analysis of the b quark fragmentation function in Z decays at LEP. *Eur.Phys.J.*, C29:463–478, 2003.

[9] Koya Abe et al. Measurement of the b quark fragmentation function in Z0 decays. *Phys.Rev.*, D65:092006, 2002.

[10] Andy Buckley, Hendrik Hoeth, Heiko Lacker, Holger Schulz, and Jan Eike von Seggern. Systematic event generator tuning for the LHC. *Eur. Phys. J. C*, 65:331–357, 2010.

[11] M. G. Bowler. lt;igt;elt;/igt;lt;supgt;+lt;/supgt;lt;igt;elt;/igt;lt;supgt;âĹŠlt;/supgt; production of heavy quarks in the string model. *Zeitschrift fÃijr Physik C Particles and Fields*, 11:169–174, 1981. 10.1007/BF01574001.

[12] T Cornelissen, M Elsing, S Fleischmann, W Liebig, E Moyse, and A Salzburger. Concepts, Design and Implementation of the ATLAS New Tracking (NEWT). *ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002*, 2007.

[13] ATLAS Collaboration.