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

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 âĂIJpunch-

[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.

throughâĂİ 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 for 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.

  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,

  $$R^{jet}_{EM} = E^{jet}_{EM}/E^{jet}_{truth} \tag{1.2}$$

  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 *i*n 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 **??** shows the final fractional jet energy scale uncertainty and its individual contributions as a function of p_T for three selected η regions. The JES uncertainty for anti-kt jets with $R = 0.4$ is between ≈4% (8%, 14%) at low jet p_T and ≈2.5%-3%

(2.5%-3.5%, 5%) for jets with $p_T > 60$ GeV in the centrak (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. **??** reflect accurately the true uncertainties in the JES.

In the case of jets induced by bottom quarks (b-jets), the jet response uncertainties are evaluated using single hadron response measurements *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.

1.2 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 [10]. 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 [11].

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 [12][13] 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.

1.2.1 Primary vertex reconstruction

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 [14].

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.

1.2.2 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[2]. On the basis that the decay point of the b-hadron must lie along its flight path, the impact parameter is signed to further discrim-

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

inate 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.

1.2.3 b-tagging algorithms

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

[11] 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.

[12] 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.

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

[14] ATLAS Collaboration.