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 it 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 algorithms for the identification of b-quark jets are presented. 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 [1], described in the previous chapter. Due to the expected level of pile-up in the LHC, one of the primary factors that influenced the decision on the jet 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 very susceptible to this effect. On the contrary, the anti-kt algorithm is the most stable after the introduction of pile-up.

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. The latter have the further advantage of being insensitive to pile-up and provide a stable reference for systematic studies.

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” [2]. Topological clusters are groups of calorimeter cells that are designed to follow the shower development taking advantage of the fine segmentataion 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 same jet algorithm as in the real data. In either case, 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 [3].

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 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 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 in-

[1]Noise-supressed towers make use of the topological clusters algorithm [2] to select cells, i.e. only calorimeter cells that are included in topo-clusters are used.

cluded 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] and 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:

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

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 [7]. 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 [8].

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 [9][10] 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 [11].

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 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.

1.2.3 b-tagging algorithms

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

Bibliography

[1] ATLAS Collaboration.

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

[3] ATLAS Collaboration.

[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-Versill, F Hubaut, R Lafaye, W Lampl, J Lvque, H Ma, E Monnier, J Parsons, P Pralavorio, Ph Schwemling, L Serin, B Trocm, 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. Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. *ATLAS-CONF-2011-102*, 2011.

[8] 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.

[9] 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.

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

[11] ATLAS Collaboration.