Identification of double b-hadron jets from gluon-splitting with the ATLAS Detector

María Laura González Silva

Doctoral Thesis in Physics

Physics Department

University of Buenos Aires

November 2012

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

Identificación de jets con hadrones *b* producidos por desdoblamiento de gluones con el detector ATLAS.

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
Lugar de Trabajo: Departamento de Física

Buenos Aires, Noviembre 2012

Agradecimientos

Quiero agradecer a mi director, Ricardo Piegaia, por darme la oportunidad de trabajar en el proyecto ATLAS, por su dedicación y su enseñanza constante; y a mis compañeros de grupo, Gastón Romeo, Gustavo Otero y Garzón, Hernán Reisin y Sabrina Sacerdoti por el trabajo compartido y por brindarme su amistad a lo largo de estos años. Quiero agradecer a Ariel Schwartzman por darnos este análisis, por su caudal inagotable de ideas y por su generosidad y la de todo su equipo. Agradezco al Laboratorio CERN, al Experimento ATLAS, a los programas HELEN y e-Planet, al CONICET y al Fundación Exactas por hacer posible la realización de esta tesis.

Quiero agradecer el apoyo de mis compañeros de la carrera, especialmente a mis amigos Cecilia y Tomás. Quiero agradecer también a mis compañeros de grupo y oficina, Lean, Yann, Javier, Pablo, y Orel por estar siempre dispuestos a darme una mano. Quiero agradecer a mis colegas y amigos de la Universidad de La Plata, Fernando, Martín y Xabier por todos los momentos compartidos; y a los amigos que hice a lo largo de estos años en mis visitas al Laboratorio CERN, Dodo, Laura, Lucile, Bárbara, Teresa, Manouk, Alex, Olivier y Haris, por ser mi familia en la distancia.

Agradezco profundamente a mis amigos y a toda mi familia por su apoyo y aliento; y de manera especial a mamá y a Juan, por comprenderme y acompañarme en todo. A ellos les dedico esta tesis.

Identificación de jets con hadrones b producidos por desdoblamiento de gluones con el detector ATLAS.

Resumen

En esta tesis se presenta un estudio de la subestructura de jets que contienen hadrones b con el propósito de distinguir entre jets-b genuinos, donde el quark b se origina a nivel de elemento de matriz (por ejemplo, en decaimientos de top, W, o Higgs) y jets-b producidos en la lluvia partónica de QCD, por el desdoblamiento de un gluón en un quark y un antiquark b cercanos entre sí. La posibilidad de rechazar jets-b producidos por gluones es importante para reducir el fondo de QCD en análisis de física dentro del Modelo Estándar, y en la búsqueda de canales de nueva física que involucran quarks b en el estado final. A tal efecto, se diseñó una técnica de separación que explota las diferencias cinemáticas y topológicas entre ambos tipos de jets-b. Esta se basa en observables sensibles a la estructura interna de los jets, construídos a partir de trazas asociadas a éstos y combinados en un análisis de multivariable. En eventos simulados, el algoritmo rechaza 95% (50%) de jets con dos hadrones b mientras que retiene el 50% (90%) de los jets-b genuinos, aunque los valores exactos dependen de p_T, el momento transverso del jet. El método desarrollado se aplica para medir la fracción de jets con dos hadrones b en función del p_T del jet, con 4,7 fb^{1} de datos de colisiones pp a $\sqrt{s} = 7$ TeV, recogidos por el experimento ATLAS en el Gran Colisionador de Hadrones en 2011.

Palabras clave: Experimento ATLAS, Jets, Subestructura de Jets, QCD, Producción de jets b, Etiquetado de Jets b.

Identification of double b-hadron jets from gluon-splitting with the ATLAS Detector.

Abstract

This thesis presents a study of the substructure of jets containing b-hadrons with the purpose of distinguishing between "single" b-jets, where the b-quark originates at the matrix-element level of a physical process (e.g. top, W or Higgs decay) and "merged" b-jets, produced in the parton shower QCD splitting of a gluon into a collimated b quark-antiquark pair. The ability to reject b-jets from gluon splitting is important to reduce the QCD background in Standard Model analyses and in new physics searches that rely on b-quarks in the final state. A separation technique has been designed that exploits the kinematic and topological differences between both kinds of b-jets using track-based jet shape and jet substructure variables combined in a multivariate likelihood analysis. In simulated events, the algorithm rejects 95% (50%) of merged b-jets while retaining 50% (90%) of the single b-jets, although the exact values depend on p_T, the jet transverse momentum. The method developed is applied to measure the fraction of double b-hadron jets as a function of jet p_T, using 4.7 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV collected by the ATLAS experiment at the Large Hadron Collider in 2011.

Keywords: ATLAS Experiment, Jets, Jet Substructure, b-jet Production, QCD, Gluon Splitting, b-tagging.

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.

Due to QCD confinement the experimental signature of quarks and gluons are not the quarks and gluons themselves but a spray of “colorless” hadrons, that we call *jets*. 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 (SM) and Beyond the Standard Model (BSM) predictions, contain jets originating from bottom (b) quarks. The ability to identify jets containing b-hadrons, the product of the hadronization of b-quarks, 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 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 "merged" 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.

The ability to single out merged b-jets has several applications. The measurement of the QCD bottom production is one of great importance due to the correspondence between parton level production and the observed hadron level, and its potential to provide information on the b-quark parton distribution function. It is found that the largest uncertainties in the theoretical calculation of the inclusive b-jet spectrum are associated to channel known as "gluon splitting" (GSP), where a gluon from the hard scatter decays into a close-by $b\bar{b}$ pair, that a jet clustering algorithm often classifies within the same jet. An improvement in the accuracy of the theoretical predictions could be achieved by not including in the production cross-section the contribution from double b-hadron jets. In addition, efficient tagging of merged b-jets can provide an important handle to understand, estimate and/or reject b-tagged backgrounds to SM and BSM analyses at the LHC that rely on the presence of single b-jets in the final state, such as top quark physics or associated Higgs production. These processes suffer from backgrounds that can be in part removed by a merged b-jet tagger.

There are two possible strategies to attempt to identify b-jets containing two b-hadrons in hadronic collisions. One of them, implemented at the CDF experiment at Fermilab [1], relies on the direct reconstruction of the two b-decay secondary vertices. 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 for the first time 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. The method developed is then applied to measure the fraction of double b-hadron jets as a function of jet p_T, using 4.7 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV collected by the ATLAS experiment in 2011.

The thesis is organized as follows: Chapter **??** 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 are described in Chapter **??**, together with a summary of the experimental conditions during the 2011 data taking. Chapter **??** details how jet reconstruction and calibration are performed at ATLAS and describes the procedure for the identification of b-quark jets. Chapter **??** 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 its systematic uncertainties are presented in Chapter **??**. Chapter **??** details the technique used for the measurement of the fraction of double b-hadron jets in QCD b-production and the associated systematic uncertainties. Finally, chapter 2 presents a summary and conclusion.

Chapter 2

Summary and conclusions

In the course of the present thesis a new method was developed to identify b-jets containing two b-hadrons which do not arise from heavy flavour production at the hard interaction but mainly via a subsequent $g \to b\bar{b}$ branching.

The method exploits the expected kinematic differences between double b-hadron ("merged") jets and single b-jets, combining a set of discriminating variables in a multivariate classifier. The differences between single and merged jets originate in the two-subjet structure of merged jets, which tend to have higher multiplicity and larger width. Several jet shape and substructure variables accounting for these envisaged characteristics were investigated in order to obtain the best single-merged discrimination. Due to the noisy environment of the hadron collisions at the LHC track-based variables were preferred over calorimeter variables.

A likelihood ratio estimator was trained using simulated QCD events. Based on discrimination power, correlation and pile-up dependence three input variables were selected for the tagger training: the jet track multiplicity, the track-jet width and the ΔR between the axes of two k_t subjets in the jet. The peformance of the tagger in Monte Carlo events was studied in bins of the calorimeter jet p_T, achieving a rejection of merged jets of over 95% (90%) for a 50% single b-jet efficiency for jets with $p_T > 150$ GeV ($p_T > 60$ GeV). A comprehensive study of the sources of systematic uncertainties in merged b-jet rejection was performed, the most relevant being the tracking efficiency and the jet energy scale and resolution with a contribution to the uncertainty of 4%, 5% and 5%, respectively. Other sources such as pile-up or the uncertainties in the track momentum resolution and the b-jet tagging efficiency proved to be negligible.

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 tool developed was used to measure the fraction of merged b-jets in QCD b-jet production. The results obtained are in very good agreement with the theoretical prediction from a QCD parton shower simulation of pp collisions.

This tool provides a handle to investigate QCD $b\bar{b}$ production and to reduce backgrounds in Standard Model 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 $ZH \to \nu\nu b\bar{b}$). Jets containing a single b-quark or antiquark also enter in many BSM collider searches, 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.

In order to expand up the results presented here, and to make further advancements in the implementation of the tagger in physics analyses the following improvements should be made: the extension to non-isolated jets using the concept of ghost-particle matching and active area of a jet for track-to-jet association and labeling and the calibration of the tagger with data. Nontheless, the study presented in this thesis demonstrates that jet substructure variables can provide a good handle for gluon splitting identification in physics searches within ATLAS.

Bibliography

[1] CDF Collaboration. Measurements of Bottom Anti-Bottom Azimuthal Production Correlations in Proton-Antiproton Collisions at $\sqrt{s} = 1.8$ TeV. *Phys.Rev.D*, 71:38, 2005.