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, Tomás y Leandro. Quiero agradecer también a mis compañeros de grupo y oficina, Javier, Yann, 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

Fraction of double b-hadron jets in data

1.1 Maximum likelihood fits

Maximum likelihood (ML) information only parametrizes the shape of a distribution (i.e. one can determine fraction of signal events from MC fits but no number of signal events).

The extended version of the maximum likelihood approach adds an extra term allowing the estimation of the absolute number of signal/background events.

For N p.d.f.s, there are N-1 fraction coefficients that should sum to less 1. The remainder is by construction 1 minus the sum of all other coefficients.

Binned or unbinned MF fit.

In most RooFit applications it doesn't matter. Internally binned data is represented the same way as unbinned data, a ROOT TTree with the bin coordinates.

Weights are supported in unbinned datasets. But use with care. Error analysis in ML fits to weighted unbinned data can be complicated...

Unbinned fits

Fitting and likelihood minimization

What happens when you do pdf->fitTo(*data)?

- Construct object representing -log of (extended) likelihood
- Minimize likelihood w.r.t floating parameters using MINUIT.

1.2 Results

True fractions predicted by PYTHIA Monte Carlo generator in Fig. 1.1

5 and 3 free parameter fits were performed. Given that single-b (mereged-b) and single-c (merged-d) templates are hard to distinguish, we fixed single-c (merged-c) fractions to single-b (merged-b) ones.

The results of the 3-parameter fits for all bins of p_T are shown in table 1.1.

The fit results are shown in Figures 1.2 and 1.3.

1.3 Systematic uncertainties

- TO DO: HACER JER SI PUEDO

In order to study the systematic uncertainties in the results the following contributions were evaluated:

- uncertainty in the track reconstruction efficiency;
- uncertainty in the jet transverse momentum resolution;
- uncertainty in the jet energy scale.

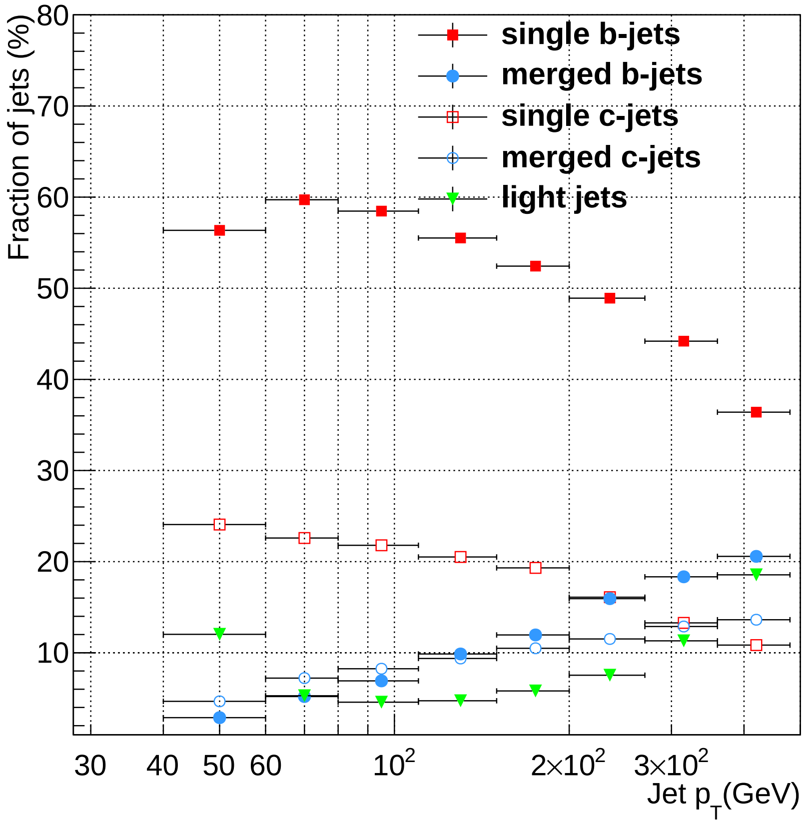

Figure 1.1: Pythia predictions of the fractions of tagged single-b, merged-b, single-c, merged-c-jets and light jets in a Monte Carlo dijet sample.

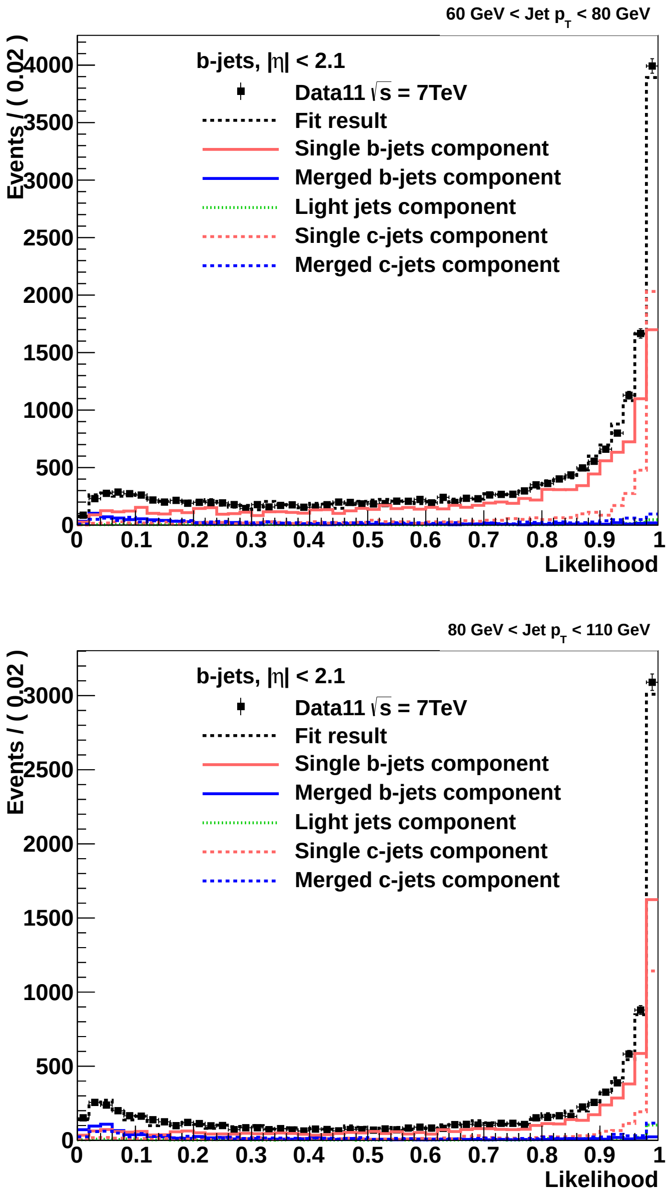

Figure 1.2: Monte Carlo templates for single, merged and light jets, fitted to data for jets between 60 GeV to 80 GeV and 80 GeV to 110 GeV.

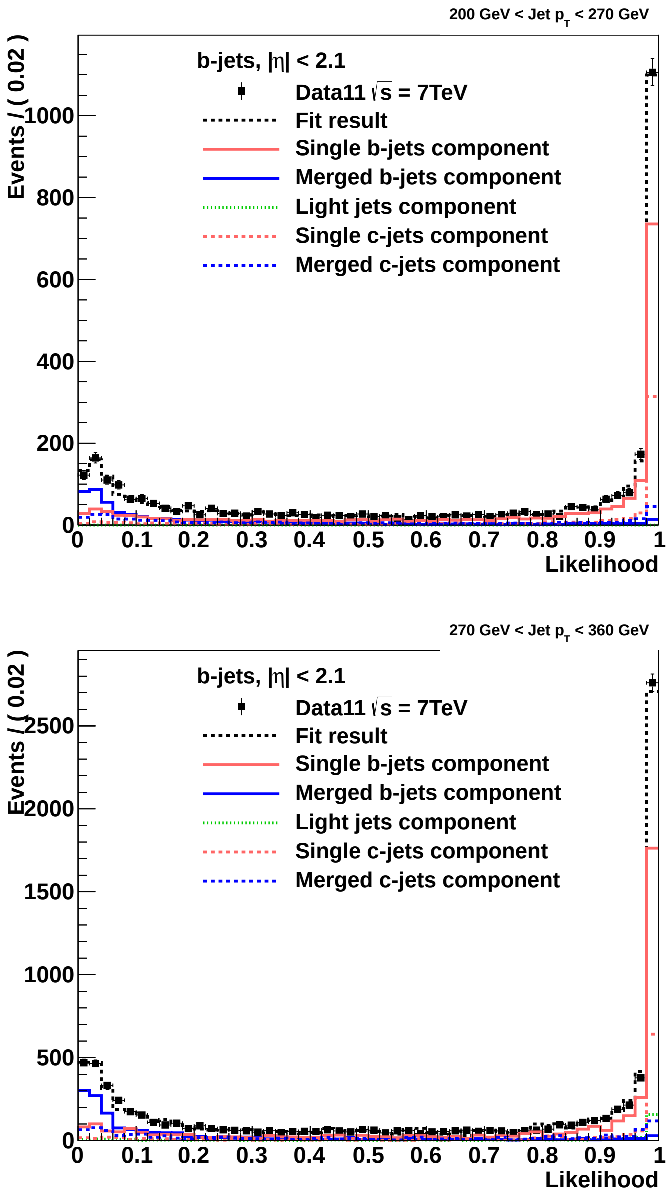

Figure 1.3: Monte Carlo templates for single, merged and light jets, fitted to data for jets between 200 GeV to 270 GeV and 270 GeV to 360 GeV.

Jet p_T (GeV)	single b-jet		merged b-jet		light jet	
	fit result	stat.err.	fit result	stat.err.	fit result	stat.err.
40 - 60	62%	3%	3%	1%	4%	4%
60 - 80	62%	1%	5.2%	0.4%	2%	2%
80 - 110	57%	1%	8.5%	0.4%	3%	2%
110 - 150	55%	2%	13%	1%	1%	4%
150 - 200	53%	3%	15%	1%	0%	4%
200 - 270	53%	5%	17%	1%	-1%	7%
270 - 360	48%	3%	19%	1%	4%	4%
360 - 480	39%	5%	21%	1%	15%	6%

Table 1.1: Measured fractions of single, merged and light b-tagged jets in experimental data from 2011 run.

The different contributions to the systematic uncertainty are summarized in Table 1.2.

Systematic source	Uncertainty
track reconstruction efficiency	negligible%
jet p_T resolution	2%
jet energy scale	2%

Table 1.2: Systematic uncertainties.

Changing the fractions merged-c/merged-b in 20% only produced a marginal effect on the fit results. The total number of merged-c plus merged-b did not changed showing that in reality we are measuring the fraction of merged b+c together. The same result is expected if changing the single-c/single-b fraction.

1.4 Enriched samples in single and merged *b*-jets

Enriched sample in merged *b*-jets

Enriched sample in single *b*-jets

The results of performing the fits on an data sample enriched in single *b*-jets is shown in tables 1.3 to 1.5. The model fitted to the data agrees well within statistics and the result is in agreement with the predictions made by PYTHIA on a sample with the same level o of enrichment.

The fit results are shown in Figures 1.4 and Figures 1.5.

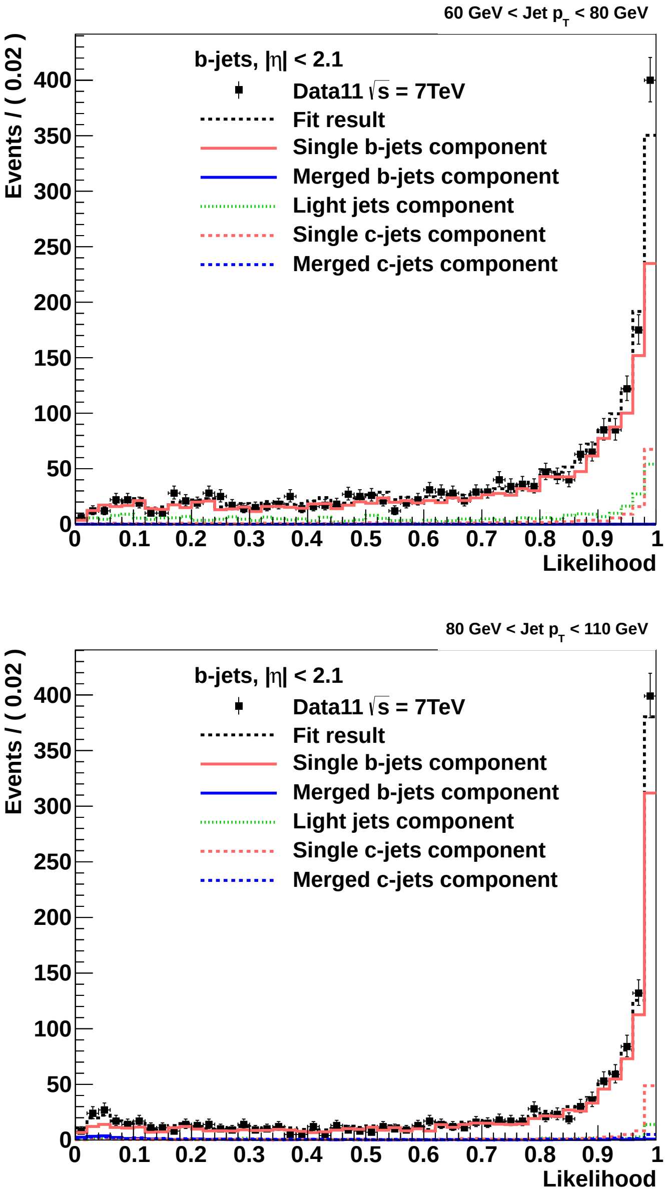

Figure 1.4: Monte Carlo templates for single, merged and light jets, fitted to data enriched in single b-jets, for jets between 60 GeV to 80 GeV and 80 GeV to 110 GeV.

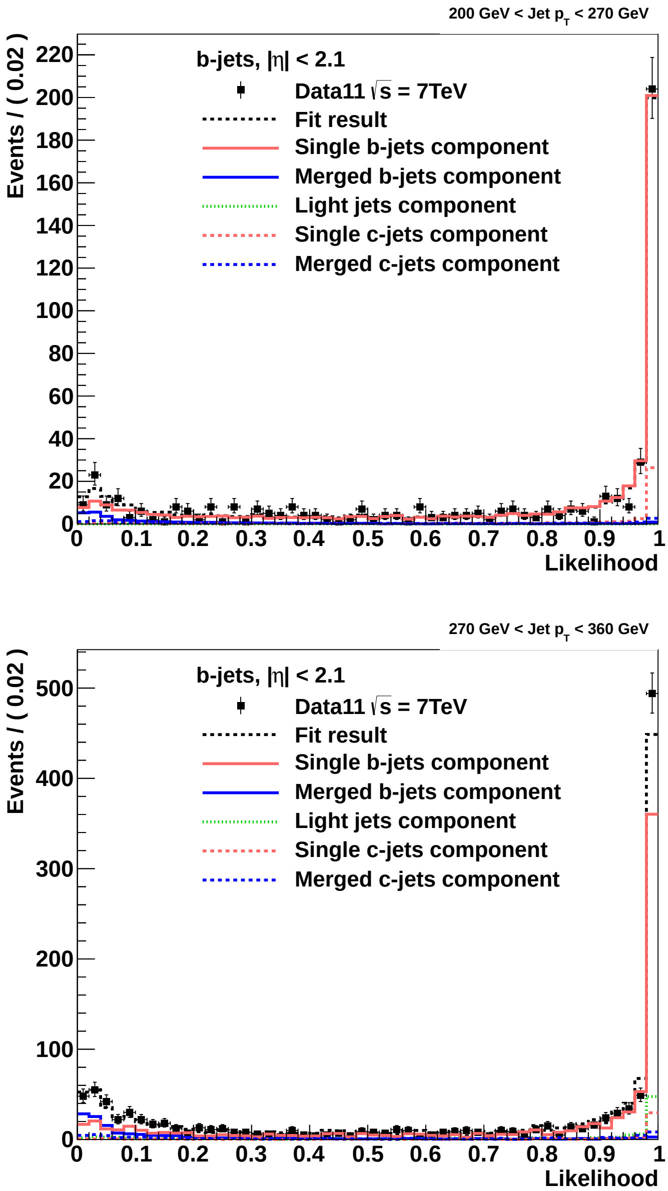

Figure 1.5: Monte Carlo templates for single, merged and light jets, fitted to data enriched in single b-jets, for jets between 200 GeV to 270 GeV and 270 GeV to 360 GeV.

Jet p_T (GeV)	single b-jet		
	fit result	stat.err.	pythia prediction
40 - 60	99%	11%	84%
60 - 80	82%	5%	87%
80 - 110	84%	5%	88%
110 - 150	86%	8%	85%
150 - 200	89%	9%	83%
200 - 270	95%	15%	80%
270 - 360	67%	11%	81%
360 - 480	73%	16%	73%

Table 1.3: Measured fractions of single b-jets in experimental data from 2011 run, enriched in single b-jets.

Jet p_T (GeV)	merged b-jet		
	fit result	stat.err.	pythia prediction
40 - 60	-1%	1%	1%
60 - 80	-3%	1%	1%
80 - 110	2%	1%	1%
110 - 150	4%	2%	3%
150 - 200	4%	2%	3%
200 - 270	7%	2%	5%
270 - 360	12%	2%	6%
360 - 480	10%	1%	8%

Table 1.4: Measured fractions of merged b-jets in experimental data from 2011 run, enriched in single b-jets.

Jet p_T (GeV)	light b-jet		
	fit result	stat.err.	pythia prediction
40 - 60	-7%	11%	5%
60 - 80	17%	6%	2%
80 - 110	4%	6%	1%
110 - 150	-1%	9%	1%
150 - 200	-6%	10%	2%
200 - 270	-17%	17%	3%
270 - 360	9%	11%	4%
360 - 480	4%	16%	8%

Table 1.5: Measured fractions of light b-jets in experimental data from 2011 run, enriched in single b-jets.

Bibliography