

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

TEMA: Double B -hadron Jet Tagging and Identification of Gluon to $b\bar{b}$ jets
with the ATLAS Detector

ALUMNA: Lic. María Laura González Silva

LU N° : 217/01

LUGAR DE TRABAJO: Dpto. de Física, FCEyN, UBA

DIRECTOR DEL TRABAJO: Dr. Ricardo Piegaia

FECHA DE INICIACIÓN: Abril de 2007

FECHA DE FINALIZACIÓN: Noviembre de 2012

FECHA DE EXAMEN:

INFORME FINAL APROBADO POR:

Alumna: Lic. María Laura González Silva

Jurado 1

Director: Dr. Ricardo Piegaia

Jurado 2

Sra. Marta Pedernera

Jurado 3

Dedicado a Cristina Silva

AGRADECIMIENTOS

A Ricardo, a mi familia...

Abstract

El detector ATLAS, uno de los cuatro experimentos del acelerador LHC actualmente en construcción en el CERN, tiene como propósito principal el descubrimiento del bosón de Higgs, la partícula involucrada en el mecanismo de ruptura de la simetría electrodébil y el origen de la masa, así como la búsqueda de nueva física más allá Modelo Estándar hasta la escala de 1 TeV. En el colisionador chocarán haces de protones con una energía de centro de masa de 14 TeV y una frecuencia de interacción de 10^9 Hz.

This thesis describes a method that allows the identification of merged b -quark jets originating from gluon-splitting. The technique exploits the kinematic differences between merged 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	Introduction	3
1.1	Identification of b -jets from gluon splitting	3
1.2	Physics Motivation	4
2	The theory of the strong interactions	9
2.1	The parton model	9
2.2	Jet physics	9
2.2.1	Monte Carlo simulations	9
2.2.2	Jet algorithms	9
3	The Large Hadron Collider	10
3.1	Luminosity and pile-up	11
4	The ATLAS Detector	12
4.1	Detector overview	12
4.2	The Inner Detector	13
4.3	The Calorimeters	15
4.3.1	El calorímetro	15
4.4	The Muon System	17
4.5	Trigger and Data Acquisition	19
4.6	Arquitectura general	19
4.7	El Nivel 1	20
4.8	El HLT	21
4.8.1	El Nivel 2	21
4.8.2	El Filtro de Eventos	22
4.8.3	El software de selección	23
4.9	Data quality	24
4.10	Simulation of particle interactions in the ATLAS Detector	24
5	Object reconstruction and event selection	25
5.1	Data and Monte Carlo samples	25

5.2	Reconstruction	26
5.3	b -jet Tagging	27
5.4	Event and jet selections	27
5.5	Track selection	29
6	BB-jet tagging	30
6.1	Introduction	30
6.2	Kinematic differences between single b - and merged $b\bar{b}$ -jets . .	30
6.3	Validation of the jet variables in data	35
6.4	Multivariate Analysis	38
6.5	Systematic uncertainties	42
6.6	Other Monte Carlo generators	45
7	Fraction of gluon-splitting jets in data	48
7.1	Template fits	48
7.2	Systematic uncertainties	48
8	Conclusions	49

Chapter 1

Introduction

1.1 Identification of b -jets from gluon splitting

The ability to identify jets containing B -hadrons is important for the high- p_T physics program of the ATLAS experiment. Two robust b -tagging algorithms taking advantage of B -hadron pair via gluon splitting, which we will henceforth call “merged” b -jets or $(b\bar{b})$ -jets, as depicted in Fig. 1.1. The ability to single out b -tagged jets from gluon splitting has several applications: measurement of QCD beauty production, $t\bar{t}$ and single top production, reduction of background in searches with b -quarks in the final state, and the study of substructure in fat jets, where $g \rightarrow b\bar{b}$ jets compete with boosted $Z \rightarrow b\bar{b}$ and $H \rightarrow b\bar{b}$ jets.

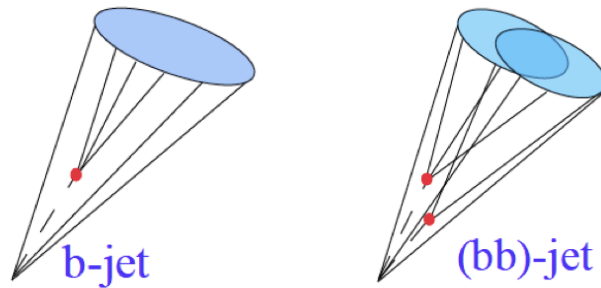


Figure 1.1: b -tagging algorithms select jets originating both from the fragmentation of a single b quark (“single” b -jets, left image) or from the splitting of a gluon into a pair of close-by $b\bar{b}$ quarks (“merged” b -jets, right image).

This thesis presents a method that allows the separation of merged and single b -jets with a multi-variate analysis that exploits their substructure

differences. The note is organized as follows. In section 1.2 we review the physics cases where this tool finds natural applications. Sections 5.1 and 5.2 review the Data and Monte Carlo samples and their reconstruction, section 5.3 discusses the b -tagging algorithms and sections 5.4 and 5.5 the criteria for selection of events, jets and tracks. The kinematic variables that differentiate between single and merged b -jets are discussed in section ?? and the validation of their MC distributions with QCD data in section ?. The construction of the multi-variate discriminator is presented in section ?? and the discussion of the systematic uncertainties in section ?. Section 6.6 investigates the performance of the tagger with other Monte Carlo generators and, finally, section ?? summarizes the results and discusses future improvements and new ideas.

1.2 Physics Motivation

Within the Standard Model (SM) a range of production channels exist for heavy-quark jets, pure QCD production or in association with heavy bosons (W, Z, H). Furthermore, b -quarks enter in many collider searches, notably because they are produced in the decays of various SM particles, top quarks and the Higgs boson (if light), and of numerous particles appearing in proposed extensions of the SM. The ability to distinguish genuine b -quark jets from those produced via gluon splitting is thus of wide application. Here we briefly discuss three cases, the measurement of QCD b quark production, the reduction of background in SM and BSM analyses with b quarks in the final state, and studies of jet substructure.

The measurement of the inclusive b -jet spectrum

The simplest and most fundamental measurement of heavy-quark jet production is the inclusive heavy-quark jet spectrum, which is dominated by pure QCD contributions. Studies of QCD bottom jets production are of intrinsic interest because of the correspondence between parton level production and the observed hadron level, and their potential to provide information on the b -quark parton distribution function, the only component of the proton structure thought to be generated entirely perturbatively from the DGLAP evolution of the other flavours. The theoretical calculation of the inclusive b -jet spectrum presents the striking feature of having rather important uncertainties ($\sim 50\%$), considerably larger than the corresponding ones for the normal (light) jet inclusive spectrum ($\sim 10\text{-}20\%$), see for example [4].

The origin of these uncertainties are reviewed in a recent paper by Banfi,

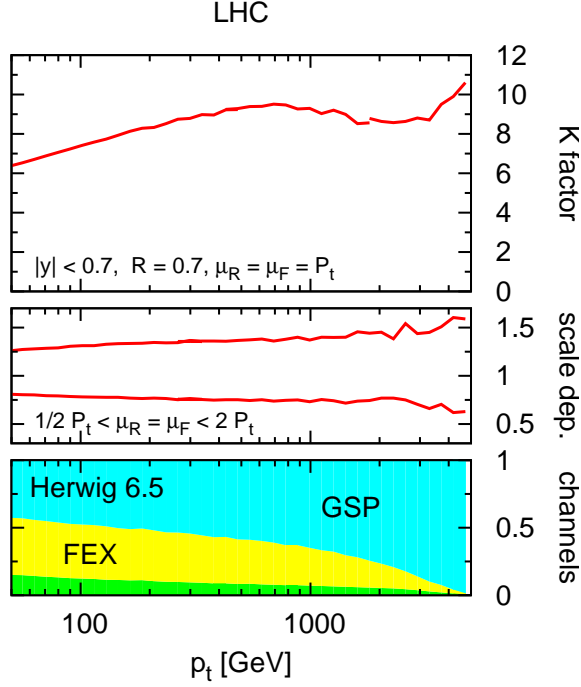


Figure 1.2: Top: K -factor for inclusive b -jet spectrum taken from [5], clustering particles into jets using the k_t jet-algorithm [6] with $R=0.7$, and selecting jets in the central rapidity region ($|y| < 0.7$). Middle: scale dependence obtained by simultaneously varying the renormalisation and factorisation scales by a factor two around p_T , the transverse momentum of the hardest jet in the event. Bottom: breakdown of the Herwig [7] inclusive b -jet spectrum into the three major underlying channels, flavor creation (FCR) flavor excitation (FEX) and gluon splitting (GSP).

Salam and Zanderighi [5], from which we have taken Fig. 1.2. Its top panel shows the K -factor, the ratio of the next-to-leading order (NLO) to the leading order (LO) cross section, obtained with MCFM for the LHC design energy (pp , $\sqrt{s} = 14$ TeV). The fact that NLO terms are considerably larger than the LO ones indicates that the perturbative series is very poorly convergent, and implies that the NLO result cannot be an accurate approximation to the full result. It is for this reason that the scale dependence (middle panels) is large. The poor convergence of the perturbative series is related to the different channels for heavy quark production. At leading order only the so-called flavour creation channel (FCR) is present, $\ell\ell \rightarrow b$, where ℓ is a generic light parton (quark or gluon), see fig. 1.3. At NLO, two new channels open

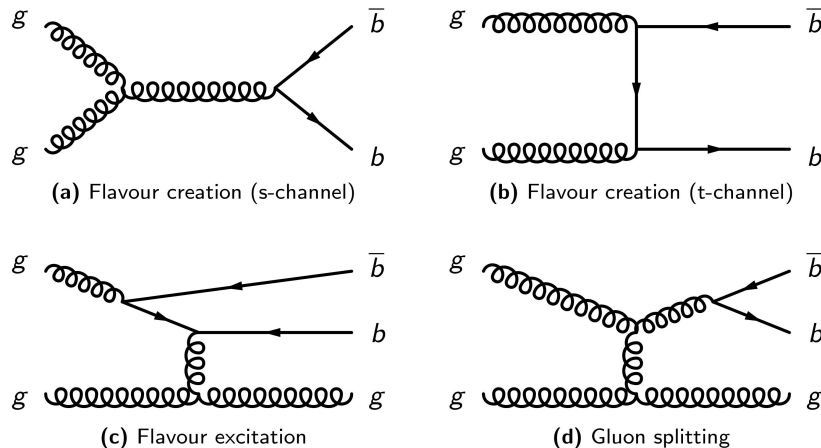


Figure 1.3: Typical Feynman diagrams for the 3 processes contributing up to next-to-leading order to QCD bottom production: (a-b) Flavor creation (FCR), (c) Flavor excitation (FEX), (d) Gluon splitting (GSP).

up, often referred to as flavour excitation (FEX) and gluon splitting (GSP). In the former, a gluon from one of the incoming hadrons splits collinearly into a $b\bar{b}$ -pair and one of those b -quarks enters the hard $b\ell \rightarrow b\ell$ scattering. In the gluon splitting process, the hard scattering LO diagram is of the form $\ell\ell \rightarrow \ell\ell$, and one of the final-state light partons (at NLO always a gluon) splits collinearly into a $b\bar{b}$ -pair that the clustering algorithm can classify within the same jet. A jet containing both b and \bar{b} is considered to be just a b -jet in standard definitions. The various channels can be conveniently separated with a parton shower Monte Carlo generator such as Herwig [7], where one can determine the underlying hard channel from the hard process in the event record. Their relative contributions to the total b -jet spectrum are shown in the bottom panel of fig. 1.2. One sees that the supposedly LO channel (FCR) is nearly always smaller than the two channels that enter only at NLO (FEX and GSP).

The largest residual uncertainties are associated with the channel with the most logarithms, gluon splitting. This channel however does not even correspond to one's physical idea of a b -jet, the one induced by a hard b -quark, and it seems somehow unnatural to include it at all as part of one's b -jet spectrum. Reference [5] thus proposes a new approach to improving the accuracy of the prediction of the b -jet spectrum, where b -jets definition maintains the correspondence between partonic flavour and jet flavour. Specifically, a jet containing equal number of b -quarks and b -antiquarks is considered to be a light jet, so that jets identified as gluon splitting are removed from the b -jet

spectrum.

Rejection of background in SM analyses and beyond-SM searches

Efficient tagging of merged b -jets from gluon splitting can provide an important handle to understand, estimate and/or reject b -tag backgrounds to Standard Model and new physics searches at the LHC.

Standard Model physics analyses that rely on the presence of b quarks in the final state such as top quark physics, either in the $t\bar{t}$ or the single top channels, and associated Higgs production: $WH \rightarrow \ell\nu b\bar{b}$ and $ZH \rightarrow \nu\nu b\bar{b}$, suffer from reducible backgrounds from QCD (that can produce b -jets from gluon splitting) and, most importantly, from the irreducible background due to W bosons produced in association with b quarks. Fig. 1.4 shows the two leading order processes that give rise to W bosons with at least one b -jet. In the first process, which can be thought of as a higher order correction to $W + \text{jets}$ production, the b quark pair is produced at small angles by gluon splitting and can often be reconstructed as a merged jet. NLO calculations

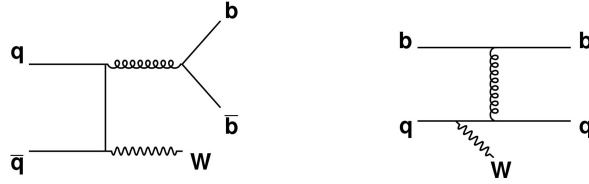


Figure 1.4: Leading order Feynman diagrams for W production in association with b quarks.

of the production of W bosons and two jets with at least one b quark at the LHC for jet $p_T > 25$ GeV, and $|\eta| < 2.5$ [8] indicate that the cross section for $W(b\bar{b})j$ is almost a factor of two higher than $Wb\bar{b}$, and about a third of Wbj , where $W(b\bar{b})j$ denotes the case in which the two b quarks are merged into the same jet.

New physics searches with b quarks in the final state can also greatly benefit from rejection of QCD and $W + b$ backgrounds with b -jets arising from gluon splitting. For example consider the search for supersymmetry in the $+ b$ -jets channel [9]. Within the framework of a generic R -parity conserving minimal supersymmetric extension of the SM The coloured superpartners of quarks and gluons, the squarks and gluinos, are expected to be copiously produced via the strong interaction at the LHC. The partners of the right-handed and left-handed quarks, q_R and q_L , can mix to form two mass eigenstates, and these mixing effects being proportional to the corre-

sponding fermion masses, they are expected to become most important for the third generation to yield sbottom (b_1) and stop (t_1) mass eigenstates significantly lighter than other squarks. Both sbottom and stop chain decay to b quarks and the lightest supersymmetric particle, producing the expected signal of $\chi + b$ -jets.

Jet substructure and boosted objects

At the LHC, many of the particles considered to be heavy at previous accelerators will be frequently produced with a transverse momentum greatly exceeding their rest mass. Good examples are the electro-weak gauge bosons W^\pm and Z^0 , the top quark, the Higgs boson or bosons and possibly other new particles in the same mass range. These boosted objects, produced either because they recoil against other energetic objects or because they arise from decays of even heavier BSM particles, can form upon decay a highly collimated topology too close to be resolved by a jet algorithm. For these cases, sophisticated tools have been developed in the last years [10, 11] to analyse the substructure of the ensuing jet and reveal its heavy-particle origin.

The study of $b\bar{b}$ jets from gluon splitting is an ideal testbed for studying jet substructure in data, as it provides a large supply of boosted, merged jets. Furthermore, understanding $g \rightarrow b\bar{b}$ jets is important as they are themselves the background to boosted object searches, like $Z \rightarrow b\bar{b}$ or $H \rightarrow b\bar{b}$. In particular, it has recently been suggested [12] that WH and ZH production become potential discovery and analysis channels by restricting ones attention to the $\sim 5\%$ of events in which the vector and Higgs bosons have large transverse momentum, $p_{TH} > 200$ GeV. Understanding the much more common QCD events with merged $b\bar{b}$ jets will be essential before attempting to measure these rare final states.

Chapter 2

The theory of the strong interactions

2.1 The parton model

2.2 Jet physics

2.2.1 Monte Carlo simulations

2.2.2 Jet algorithms

Chapter 3

The Large Hadron Collider

El Large Hadron Collider (LHC) es un acelerador de partículas actualmente en su fase final de construcción, localizado en el Centro Europeo de Investigación en Física de Partículas (CERN) en Ginebra, Suiza. Se prevé que comience su operación en el año 2008. Este acelerador usa el túnel de 27 km de circunferencia creado para el colisionador Large Electron-Positron (LEP) y a diferencia de éste, en el nuevo se producirán colisiones protón-protón, con una energía de centro de masa de 14 TeV (7 veces mayor a los 2 TeV alcanzados por el acelerador Tevatron en el laboratorio Fermilab en Chicago, Estados Unidos). El LHC también podrá emplearse para hacer colisionar iones pesados; la colisión en este caso tendrá una energía de 1150 TeV.

El sistema de inyección del LHC se esquematiza en la figura 3.1. Los protones en el LHC serán producidos por un acelerador lineal de protones de 50 MeV, que será inyectado luego en el Proton Synchrotron Booster de 1.4 GeV. El Proton Synchrotron (PS) acelerará los protones hasta alcanzar los 25 GeV y generará grupos (*bunches*) de $\sim 10^{11}$ protones cada 25ns. El Super Proton Synchrotron (SPS) acelerará los protones hasta los 450 GeV, para ser inyectados luego dentro del anillo del LHC[?].

Figure 3.1: EL anillo del LHC y su sistema de inyección

El campo magnético necesario para mantener el haz de protones circulando en la máquina es provisto por 1232 dipolos superconductores, dando lugar a un campo de 8.4 Tesla. Los *bunches* de protones separados por 25 ns y con una longitud media de 7 mm se cruzan en cuatro puntos del acelerador. En estos puntos se encuentran situados los detectores de los experimentos del LHC: ATLAS y CMS, experimentos de propósito general diseñados para la

búsqueda de nueva física y mediciones de precisión; LHCb, dedicado a estudiar la violación de CP y la física de mesones B y ALICE, que estudiará la física de iones pesados y el comportamiento de la materia nuclear a muy altas energías y densidades.

En cuanto al próximo funcionamiento del LHC, dos fases están previstas: en los primeros años de operación (fase de baja luminosidad) la luminosidad nominal esperada es de $2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$, para luego alcanzar un valor de $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ (fase de alta luminosidad).

3.1 Luminosity and pile-up

Chapter 4

The ATLAS Detector

4.1 Detector overview

El detector ATLAS, acrónimo para A Toroidal LHC Apparatus, fue diseñado para estudiar la física de colisiones $p-p$ en el LHC. Todo el detector está contenido en un cilindro de aproximadamente 44 metros de largo por 22 metros de diámetro y pesa unas 7000 toneladas[13]. Presenta una estructura tipo cebolla, pudiendo separarse en tres sistemas comenzando desde el punto de interacción: el detector interno, los calorímetros y el sistema de muones (figura 4.1) . Cada parte se divide a su vez en más capas. Las partículas emergentes atraviesan primero el detector interno, donde se reconstruye la trayectoria de las partículas cargadas. En el calorímetro todas las partículas, con excepción de muones y neutrinos, depositan toda su energía y se detienen. Los muones interactúan electromagnéticamente al igual que los electrones, pero al ser mucho más masivos que éstos no emiten radiación de frenado y logran atravesar el calorímetro dando señal en el sistema de detección de muones. ATLAS cuenta además con un sistema de imanes que provee de un campo magnético que curva las trayectorias de las partículas cargadas que atraviesan el detector interno y el espectrómetro de muones, permitiendo medir su momento.

Todos los detectores de ATLAS, así como el sistema de imanes, se componen de un barril central y dos tapas laterales idénticas. La cobertura en pseudorapidez de cada una de estas partes dependerá del sistema considerado.

Figure 4.1: El detector de ATLAS

4.2 The Inner Detector

El detector interno de ATLAS (figura 4.2) está contenido en un cilindro de 7 m de longitud y 1,15 m de radio exterior. El mismo ha sido diseñado para la reconstrucción de trazas de partículas cargadas en un campo magnético solenoidal de 2 Tesla, con un rango de pseudorapidez que se extiende hasta $|\eta| = 2.5$. Está compuesto por tres subdetectores: el detector de píxeles, el detector de microbandas de silicio (SCT, del inglés *Semiconductor Tracker*) y el de transición de radiación (TRT, *Transition Radiation Tracker*). La combinación de estas tecnologías permite el reconocimiento de trazas de manera robusta y una alta precisión en las coordenadas η y ϕ .

Desde el punto de vista mecánico el detector consiste en un barril extendido en la coordenada z sobre ± 80 cm, y dos ruedas o tapas laterales idénticas cubriendo el resto de la cavidad. En la región del barril todos los elementos de detección están ordenados en estructuras cilíndricas, mientras que en las tapas dichos elementos están montados en discos perpendiculares a la dirección del haz. Esto asegura que las partículas pasen todos los elementos de detección con ángulos de incidencia grandes.

Figure 4.2: Vista del detector interno de ATLAS. Se indican las posiciones del barril central y partes laterales del SCT, el detector de píxeles y el TRT

El detector de píxeles

Este sistema consiste en tres barriles de ~ 4 , 10 y 13 centímetros de radio medio, respectivamente; y 5 discos a cada lado. Contiene 140 millones de elementos detectores de forma cuadrada, cada uno de $50 \mu\text{m}$ en la dirección $R\phi$ y $300 \mu\text{m}$ en z , midiendo así dos coordenadas por cada módulo detector. Todo el dispositivo está situado tan cerca como es posible del punto de interacción, entregando 3 mediciones de alta precisión y granularidad en la región cercana al punto de interacción primario, contribuyendo a la medición del parámetro de impacto y la posición del vértice.

El detector de microbandas de silicio

El sistema de detección por microbandas de silicio consta de cuatro cilindros concéntricos; de 300, 373, 447 y 520 mm de radio, respectivamente, compuestos, cada uno de ellos, por módulos detectores montados en estructuras

de fibra de carbón. Cada módulo posee dos capas de microbandas. En la primera, las bandas están dispuestas paralelas al eje del haz midiendo así la coordenada ϕ directamente. Éstas, junto con las microbandas en la capa siguiente, con un ángulo *stereo* de 40 mrad, reconstruyen la coordenada z . A los lados del barril central se tienen dos tapas con módulos montados en 9 ruedas. La resolución espacial en el sistema es de 16 μm en $R\phi$ y 580 μm en z .

Tendremos, típicamente, 8 capas de bandas de silicio (más las tres de píxeles) atravesadas por cada traza proveniente del punto de interacción. Un impacto (o *hit*) en una de las capas se referirá a un canal de detección con una señal de salida sobre un determinado umbral de energía. En el caso del detector de silicio, tendremos un cluster pequeño de canales, creado por una partícula cargada o ruido. Dos impactos correspondientes a una misma traza, en las dos capas de un módulo detector contribuyen a un punto tridimensional en el espacio que llamaremos SP, del inglés *Space Point*.

El detector de radiación de transición

El detector de radiación de transición de ATLAS está basado en el uso de tubos detectores del diámetro de una pajita de gaseosa, para la identificación y reconstrucción de las trayectorias de partículas cargadas. En total hay alrededor de 370000 tubos de 4 mm de diámetro y 144 de longitud en todo el detector, y, en virtud de su pequeño diámetro pueden operar a las altas frecuencias esperadas en el LHC.

Cada tubo contiene en su interior un hilo de tungsteno, cubierto de oro, de 30 μm de diámetro y está relleno de una mezcla gaseosa (70 % xenón) que se ioniza cuando pasa una partícula. La medición del tiempo de arribo de los iones producidos al hilo central, permite determinar la posición del impacto dentro del radio del tubo. Los tubos detectores están orientados de manera radial (en ruedas) en las tapas y, a lo largo del eje del haz, en la región del barril. Estas orientaciones han sido elegidas para maximizar el número de tubos atravesados en todas las direcciones. La sección del barril consiste en módulos individuales de entre 329 y 793 tubos axiales, cubriendo el rango radial de 56 a 107 cm.

La detección de radiación de transición, provocada por la existencia de materiales de distinto índice de refracción entre los tubos, generará señales más intensas y permitirá una mejor identificación de electrones que atraviesan el detector (los muones también emiten radiación de transición pero al ser mucho más masivos que los electrones, lo hacen en una cantidad mucho menor).

Este sistema de detección por tubos provee un gran número de puntos por

traza (típicamente 36 puntos), lo cual determina un seguimiento continuo de la misma, con mucho menos material por punto y menor costo (mucho menor comparado con la tecnología de píxeles implementada en el subdetector más interno). La gran cantidad de puntos o impactos por traza es un instrumento poderoso en la búsqueda y reconstrucción de trazas en el detector interno.

4.3 The Calorimeters

El sistema de imanes superconductores de ATLAS consiste en un solenoide central que provee el campo magnético necesario al detector interno, rodeado por un arreglo de bobinas o bucles, con forma de pista de carrera, que generan un campo magnético toroidal para el espectrómetro de muones. Todo el sistema es enfriado de manera indirecta mediante el flujo de helio líquido a 4,5 K.

El solenoide es un electroimán superconductor de 5,3 m de largo, situado en el interior del calorímetro electromagnético. Comparte el crióstato con el calorímetro de argón líquido, evitando la presencia de dos paredes criostáticas y reduciendo así la cantidad de material introducido. La longitud del solenoide es considerablemente más pequeña que la del barril del detector de trazas. Este es el resultado de un compromiso: un bobinado corto reduce la cantidad de material introducido mientras que uno largo proporciona un campo magnético más uniforme en dicho detector. El campo magnético a lo largo del eje z es de 2 T en el punto de interacción.

El arreglo de bobinas está dividido en un barril central y dos regiones laterales, al igual que los detectores. El barril central está constituido por 8 bobinas de 5 m de ancho por 25 metros de largo aproximadamente, dispuestas simétricamente alrededor del haz de manera radial. Las bobinas del barril se encuentran en crióstatos separados, mientras que las 8 bobinas en cada una de las tapas o toroides laterales están ubicadas en un crióstato común.

Con un campo magnético toroidal las partículas atravesarán todo el rango de pseudorapidez casi perpendicularmente al haz. El número pequeño de bobinas que generan el campo toroidal resulta en una intensidad de campo que varía fuertemente con la coordenada ϕ . En el barril el campo magnético es de 2 T, mientras que en las tapas es de 4 T en las zonas de mayor intensidad.

4.3.1 El calorímetro

Una vez que la partícula atraviesa el detector interno, ingresa al calorímetro. Un detector calorimétrico está diseñado para absorber la energía de las

partículas que lo atraviesan y se encuentra dividido normalmente en un calorímetro electromagnético y uno hadrónico dado el diferente comportamiento de fotones/electrones por un lado, y hadrones por el otro. Así el calorímetro de ATLAS consiste en un sector electromagnético que cubre una región de pseudorapidez $|\eta| < 3.2$ (barril y tapas), y uno hadrónico formado por tres partes: un gran barril cubriendo la zona de $|\eta| < 1.7$, dos tapas en los extremos cubriendo la región de $1.5 < |\eta| < 3.2$, y calorímetros de bajo ángulo cubriendo la región de pseudorapidez de $3.1 < |\eta| < 4.9$.

El calorímetro de argón líquido

El calorímetro electromagnético de ATLAS consiste en un gran barril interno y dos ruedas (*end-caps*), una a cada lado del mismo. Se trata de un calorímetro de muestreo que funciona con argón líquido (se denomina de muestreo porque mide sólo una fracción de la energía depositada por la partícula incidente). Está formado por placas de plomo/acero inoxidable intercaladas con electrodos de cobre; el espacio entre ambos se llena con argón líquido, permitiendo la deriva de los electrones de ionización bajo el alto voltaje aplicado.

Con el fin de asegurar una perfecta cobertura en todo el rango de pseudorapidez y proveer una completa simetría en la coordenada ϕ se ha elegido una geometría en forma de acordeón. En la región dedicada a física de precisión ($|\eta| < 2.5$) el calorímetro electromagnético está dividido en cuatro secciones de muestreo. La primera o *Presampler* consiste en una delgada capa de argón desprovista de material absorbente, cuyo propósito es la corrección por la pérdida de energía en el solenoide y las paredes del crióstato. La segunda sección posee una profundidad de $4.3X_0$ ¹ y en ella la lectura se lleva a cabo mediante celdas en forma de tiras delgadas en η , dando una buena resolución en dicha coordenada, con $\Delta\eta = 0.0031$. La sección de *2nd Sampling* ($16X_0$) es donde se deposita la mayoría de la energía, teniendo ambas coordenadas igual importancia. Allí el tamaño de celda es de $\Delta\phi\Delta\eta = 0.0245 \times 0.0245$. Sólo los electrones más energéticos llegarán a la cuarta sección de muestreo (*3rd Sampling*).

Las ruedas calorimétricas comienzan en $|\eta| = 1.5$ y continúan abajo hasta $|\eta| = 3.2$, pero con un tamaño de celda mayor por encima de $|\eta| = 2.5$.

En la región de bajo ángulo, el calorímetro hadrónico funciona también

¹ X_0 es el símbolo utilizado para la Longitud de Radiación (*Radiation length*), definida como la distancia media sobre la cual un electrón muy energético pierde $1/e$ de su energía por bremsstrahlung o bien, como $7/9$ del camino libre medio en la producción de pares. Constituye una escala de longitud apropiada para describir cascadas electromagnéticas de alta energía.

con argón líquido de manera de resistir los altos niveles de radiación. Su diseño es más sencillo que el del calorímetro electromagnético y como absorbentes posee placas paralelas de cobre, perpendiculares al haz. La región de esta parte del calorímetro hadrónico, que cubre hasta $|\eta| = 4.9$, está hecha de cobre/tungsteno. La elección de este material es necesaria para limitar el ancho y la profundidad de las lluvias provenientes de jets de altas energías cercanas a la línea del haz.

El calorímetro hadrónico de tejas

El calorímetro hadrónico de tejas de ATLAS está compuesto por tres barriles, uno central de 5,6m y dos extensiones de 2,9m cada una. El radio interno es de 2,2m y el externo, de 4,2m.

Cada barril está dividido en 64 cuñas azimutales o módulos, con una estructura periódica en la dirección paralela al haz. Cada módulo es una estructura de tejas de hierro (material absorbente) alternadas con tejas de plástico centellador, dispuestas en un plano paralelo al eje del haz. Los materiales centelladores emiten luz en forma de pequeños pulsos cuando son atravesados por partículas o radiación; acoplando el centellador a un fotomultiplicador el pulso de luz se convierte en un pulso eléctrico que puede ser analizado. En el caso del calorímetro hadrónico de ATLAS, cuando una partícula atraviesa una teja centelladora emite luz en el rango del ultravioleta, de intensidad proporcional a la energía depositada por la partícula.

Se ha elegido una segmentación proyectiva en torres de $\Delta\phi\Delta\eta = 0.1 \times 0.1$ y cada una de estas torres está dividida, en profundidad, en tres celdas, leídas individualmente por dos fotomultiplicadores para conseguir redundancia en la señal. La luz generada en las tejas es recogida mediante fibras ópticas que cambian la longitud de onda y transportada a los fotomultiplicadores (este subdetector posee unos 10.000 fotomultiplicadores).

El calorímetro hadrónico debe tener el espesor suficiente para contener la energía de los hadrones. Con este propósito y para obtener una buena resolución se han elegido 11 longitudes de absorción como camino previo a las cámaras de muones.

4.4 The Muon System

El sistema de muones sirve a un doble propósito: funciona como sistema de disparo (o *trigger*) para la selección de eventos con muones de alta energía, y como espectrómetro de muones de alta precisión. En este sentido, este detector llevará a cabo la identificación de los muones producidos en las col-

isiones p - p , determinando sus trayectorias y momentos. El sistema consiste en un conjunto de toroides (llamamos así, por su forma, a los tres conjuntos de bobinas que proveen el campo magnético toroidal) y cámaras de tubos de deriva que se encuentran rodeando al calorímetro. En la parte del barril del detector, las cámaras están situadas en el interior del toroide lo que permite la medición del momento de las partículas a partir de la desviación de sus trayectorias en el campo magnético. En las tapas, donde la presencia del crióstato impide posicionar las cámaras dentro del campo magnético, el momento es medido a partir de la diferencia entre los ángulos de entrada y salida del imán. En el plano transversal, tanto en la región del barril como en las tapas laterales, el sistema de cámaras estará dividido en 16 sectores, siguiendo la simetría determinada por las 8 bobinas del barril central del sistema magnético. Las cámaras cubren el espacio entre las bobinas, y todo el rango acimutal en la región que las rodea. Los sectores se numeran comenzando a partir de $\phi = 0$, en el sentido contrario de las agujas del reloj, teniendo en la dirección vertical a los sectores 6 (en la parte superior del detector) y 13 (sector inferior).

Los cámaras de tubos de deriva (MDTs) son cámaras proporcionales hechas de tubos de aluminio de 30 mm de diámetro y longitudes variables de 70 a 630 cm, con un hilo central de $50\mu\text{m}$ de diámetro, de W-Re. En la región del barril dichas cámaras están distribuidas en 3 capas cilíndricas concéntricas (estaciones) alrededor del haz, de 5; 7,5 y 10 metros de radio. Los tubos están dispuestos de manera transversal al eje z de manera de medir la coordenada en el plano de desviación de la trayectoria de la partícula (plano Rz). Estas cámaras miden el tiempo de deriva de la ionización producida por el paso del muón, teniendo una resolución de $80\mu\text{m}$.

Cada cámara MDT está cubierta por una o dos cámaras de placas resistivas (RPCs). Cada una de ellas encierra un volumen de gas entre planchas resistivas de baquelita, dotada una de ellas con tiras de electrodos. Dado que los tubos de deriva poseen un diámetro relativamente grande que resulta en un tiempo de deriva máximo de 480ns, mucho mayor que los 25 ns entre cruce de *bunches*, se requieren cámaras especiales de disparo para la selección de eventos. La función de trigger en el barril es provista por tres capas de RPCs, situadas, dos de ellas, a ambos lados de la segunda estación de MDTs y la restante, en la cara interior de la estación más externa. En las tapas, esta función es cumplida por tres estaciones de TGCs (*Thing Gap Chambers*). Estas cámaras son similares en diseño a cámaras proporcionales multihilo, con la diferencia de que poseen una distancia cátodo-cátodo menor que la pendiente del ánodo (hilo). Las cámaras de disparo proveen una estimación de las coordenadas ϕ y η del punto de impacto de la traza, mientras que las cámaras MDTs darán (con mayor precisión) la coordenada η .

En la región de bajo ángulo, donde la densidad de trazas es mayor, se utilizan cámaras de tiras de cátodos (CSCs) de granularidad más fina comparadas con las MDTs, para la detección de trayectorias. Estas cámaras son cámaras proporcionales, con un espacio entre hilo de 2,5 mm. Cada una de ellas proporciona medida de dos coordenadas y puede operar en condiciones de alto campo magnético.

En la figura 4.3 se puede ver un esquema del espectrómetro de muones, donde se indica la posición de las diferentes cámaras descriptas.

Figure 4.3: Vista tridimensional del espectrómetro de muones de ATLAS, indicando las áreas cubiertas por las diferentes cámaras que lo componen

4.5 Trigger and Data Acquisition

En este capítulo se analiza la estructura del trigger de ATLAS, y el sistema de adquisición y flujo de los datos. Se presenta, asimismo, una breve descripción de los algoritmos usados en la reconstrucción de trayectorias para la selección de eventos en el detector interno.

4.6 Arquitectura general

El sistema de Trigger y Adquisición de Datos[?] de ATLAS está basado en tres niveles de selección *online*: Nivel 1, Nivel 2 y Filtro de Eventos. Cada nivel es más lento pero más preciso que el anterior. Trabajando con una frecuencia de interacción de 10^9 Hz y luminosidades del orden de $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, este sistema será el encargado de reducir la frecuencia de eventos inicial de 40 MHz a 200Hz, que es la velocidad con la que pueden almacenarse.

En la figura 4.4 se muestra un vista simplificada de los principales componentes y funciones.

Figure 4.4: Principales componentes del sistema de trigger y adquisición de datos de ATLAS.

El mecanismo que lleva a cabo el movimiento de la información (Data Flow System), es el responsable de recibir los datos de los detectores, pasando

parte de ellos al sistema de trigger y enviando luego, los eventos seleccionados al lugar de almacenamiento. Siguiendo el esquema de la figura, la comunicación entre los *drivers* de lectura de cada detector (RODs) y el sistema de adquisición de datos, está dada por los *buffers* de almacenamiento transitorio (ROBs). La información de los eventos aceptados por el Nivel 1 son transportados de los primeros al sistema de lectura (ROS), que consta de numerosos ROBs, guardando los datos a la espera de la decisión del trigger. La información requerida por el segundo nivel es provista por estos últimos. Los eventos aceptados son reconstruidos (a partir de fragmentos contenidos en diferentes ROBs) y pasados al siguiente nivel.

El Nivel 2 y el Filtro de Eventos componen el *High-level Trigger* (HLT) de ATLAS. El Nivel 2 trabaja a la frecuencia de aceptación del Nivel 1, utilizando una secuencia de rápidos algoritmos de selección que operan típicamente sobre una fracción de los datos del evento, contenida en regiones del detector previamente seleccionadas por ese nivel (ver el mecanismo de la región de interés en la siguiente sección). Si la decisión del Nivel 2 es rechazar el evento, los datos del mismo son eliminados de los buffers correspondientes. Si el evento es aceptado, se reconstruye en el EB (Event Builder) y es pasado al Filtro de Eventos. Este nivel ejecutará algoritmos de reconstrucción más sofisticados, adaptados de aquellos para el análisis *offline*, utilizando información detallada de los detectores para efectuar el proceso de selección final, que determinará cuáles son los eventos que serán guardados para posteriores estudios.

En las siguientes secciones se presenta una descripción más detallada de los niveles de trigger.

4.7 El Nivel 1

El primer nivel de trigger de ATLAS es implementado mediante hardware. Éste realiza una decisión inicial a partir de la información provista por los calorímetros y del detector de muones, basando su estrategia en la combinación de objetos en coincidencia.

En el sistema de muones, los candidatos de alto momento transversal son identificados en las cámaras especiales de trigger: RPCs en el barril y TGCs en las tapas. En el caso del calorímetro, se definen una serie de conjuntos de umbrales de p_T para cada objeto (electrones, fotones, jets, etc.), seleccionando aquellos que pasen los criterios de selección correspondientes al evento físico de interés.

Puesto que la decisión de aceptar un evento no puede ser realizada en los 25 ns que median entre dos cruces de *bunches*, los subdetectores almacenan

localmente la información del mismo en *pipelined buffers* hasta que el Nivel 1 efectúa la selección. Luego, los datos son enviados a los RODs específicos de cada detector para luego dirigirse a los ROBs, donde son almacenados hasta que la decisión del Nivel 2 sea alcanzada. Cuando un evento es aceptado, el Nivel 1 comunica la decisión al mecanismo que se encargará de construir una Región de Interés (RoI). Este mecanismo es una importante pieza sobre la que descansa la estrategia del sistema de trigger; a través del mismo, el Nivel 2 hará uso de la información del evento en regiones localizadas del detector, de manera que los algoritmos de reconstrucción en ese nivel sólo transfieran los ROBs necesarios para arribar a una rápida decisión. La RoI contendrá la información de la posición (η y ϕ) y el momento de los objetos candidatos.

Este nivel está diseñado para llevar a cabo su decisión en un tiempo menor a $2.5 \mu\text{s}$, medidos desde la colisión p - p , hasta que la información del evento está disponible en la electrónica de salida de los detectores. En este proceso la frecuencia de eventos será reducida a 75KHz (límite fijado por la electrónica).

4.8 El HLT

El High-level Trigger de ATLAS abarca la segunda y tercera etapa de la selección de eventos. Comprende el Nivel 2 y el Filtro de Eventos, y contiene además, el Software de Selección (ESS). Este último comparte la estructura usada por el Offline para los códigos de selección, facilitando el análisis *offline* de los datos, y el desarrollo de algoritmos en el HLT.

El punto de entrada del trigger es el resultado del Nivel 1. Éste provee información acerca de la región de interés, fundamental para el rápido funcionamiento de los algoritmos del Nivel 2. Así, los datos del Nivel 1 guían la selección del Nivel 2; y ésta a su vez guiará la del Filtro de eventos, como se ilustra en la figura 4.5.

Figure 4.5: Cadena de selección del *High-level Trigger* de ATLAS. Cada nivel es guiado por el resultado del paso anterior.

4.8.1 El Nivel 2

La tarea específica del Nivel 2 es reducir la frecuencia de eventos de $\sim 100 \text{ kHz}$ a alrededor de 2 kHz , combinando la información de todos los detectores para

su decisión global. A diferencia del Nivel 1, esta segunda etapa de selección realiza operaciones no sincronizadas sobre los eventos, con un tiempo de decisión de 10 ms.

El Nivel 2 utiliza las regiones de interés provistas por el Nivel 1. Cada región es examinada en el subdetector de origen (calorímetro o sistema de muones) para su confirmación; para luego buscar información de otros subdetectores. En el caso del trigger de muones, el poder de rechazo del Nivel 2 proviene de ajustar los umbrales de p_T , respecto de los utilizados en el primer nivel, a partir de la información de las cámaras de precisión del sistema de muones (MDTs) y la correspondiente al detector interno. Los procesadores del Nivel 2 son los encargados de ejecutar luego el software de selección de eventos, utilizando la información almacenada en los *buffers*. Usando las RoIs del Nivel 1, el Nivel 2 accederá de manera selectiva a los datos en los ROBs, moviendo sólo la información requerida para efectuar la decisión. Típicamente, sólo una pequeña fracción del detector, correspondiente a las regiones centradas en los objetos indicados por el Nivel 1, serán necesitados por el segundo nivel.

Hasta que un evento es aceptado o rechazado (en ~ 10 ms), los datos son retenidos en los ROBs. En caso de aceptación, los fragmentos del evento almacenados en distintos buffers serán requeridos por el sistema de control del Nivel 2 (L2SVs) para ser enviados al constructor de eventos (EB). El evento ensamblado es guardado en una única dirección de memoria para ser utilizado por el Filtro de Eventos. El tamaño promedio de un evento será del orden de 1,5 MB.

4.8.2 El Filtro de Eventos

Luego del Nivel 2, la última etapa de selección *online* es realizada por el Filtro de Eventos (EF). El EF emplea algoritmos y métodos similares a los implementados en el análisis *offline*, adaptados para su corrida en el tiempo real del experimento; su poder de rechazo radica en el uso de algoritmos y criterios de selección más complejos, que por límites en el tiempo de procesamiento no pueden ser utilizados en el Nivel 2.

El EF utilizará información actualizada de la calibración y alineamiento del detector y un completo mapa del campo magnético; llevando a cabo con ello la selección final del evento físico que será guardado para su estudio en el Offline. La frecuencia de aceptación del nivel anterior será reducida en un orden de magnitud, almacenando a una tasa de ~ 100 MB/s.

4.8.3 El software de selección

La tarea del software de selección (ESS) es la selección y clasificación de los eventos. Candidatos tales como electrones, jets, muones, etc., representados por objetos abstractos, son reconstruidos utilizando un particular conjunto de algoritmos. Un evento es seleccionado si el objeto reconstruido satisface al menos una de las firmas establecidas en el menú del sistema de disparo. En el Nivel 2 y el Filtro de Eventos (EF), los eventos serán rechazados si no pasan los específicos criterios de selección, diseñados para la reducción de la frecuencia de eventos, al límite dado por la velocidad a la que éstos pueden ser almacenados.

El ESS se compone de una infraestructura y un conjunto de programas de selección para las dos etapas del HLT. Los algoritmos de reconstrucción para el trigger están basados en aquellos utilizados para la reconstrucción *offline*, pero correrán *online* en el entorno de software provisto por los procesadores del Nivel 2 y el EF.

De manera de facilitar el desarrollo de los algoritmos del HLT y simplificar los estudios del Offline; el ESS ha sido diseñado de manera de poder ser ejecutado directamente en el entorno provisto por la estructura de software de análisis offline del experimento, ATHENA[?]. La estructura dada por este paquete de software es lo suficientemente flexible como para abarcar una variedad de procesos, incluyendo no sólo algoritmos de trigger sino también tareas de calibración y monitoreo. Se ha destinado un apéndice (A) para su descripción.

En el Offline, la tarea del ESS es la de emular la cadena completa de selección *online*. Para su ejecución el sistema se sirve de cuatro sub-paquetes: el direccionamiento o *Steering*, los algoritmos del HLT, y los paquetes de software para la clasificación y movimiento de los datos, EDM (*Event Data Model*) y el DM (*Data Manager*). Los últimos toman los datos del evento en el formato que poseen a la salida de los sistemas de lectura (*Raw data* en formato *byte stream*), y los convierten en objetos que puedan ser usados por los algoritmos en la cadena de selección (*Raw Data Objects*).

La tarea de los algoritmos del HLT es la de analizar los datos del evento, reconstruyendo partes del mismo, luego de la selección del Nivel 1. El paquete se compone de dos subconjuntos principales:

- Programas de preparación de datos. Son los algoritmos ejecutados por los sistemas EDM y DM para la conversión del formato de los datos del evento.
- Algoritmos FEX o de *Feature Extraction*. Comprende los programas de reconstrucción y los llamados algoritmos de “hipótesis”. Estos últimos

(a los primeros nos referiremos en la siguiente sección) son aquellos programas que se encargan de eliminar, una vez realizada la reconstrucción, aquellos candidatos que no cumplen con las características o atributos asignados al evento físico en consideración (hipótesis), aplicando específicos criterios de selección. La presencia de los algoritmos de hipótesis es fundamental en la secuencia del HLT ya que evita la ejecución innecesaria de algoritmos al descartar eventos en las primeras etapas de la cadena.

Por último, el subpaquete de *Steering* es aquel que organiza el procesamiento de los datos del evento en el Nivel 2 y el Filtro de eventos; controlando el orden en el que los algoritmos de reconstrucción e hipótesis son ejecutados. El *Steering* define la secuencia del HLT, y manipula los resultados en cada paso de selección de manera que la decisión del trigger sea alcanzada.

4.9 Data quality

4.10 Simulation of particle interactions in the ATLAS Detector

Chapter 5

Object reconstruction and event selection

5.1 Data and Monte Carlo samples

The method presented in this note relies on Monte Carlo predictions for the signal (single b) or background (merged b) hypotheses, but we assess with experimental data the extent to which the simulation accurately reproduces the measured distributions for the different variables explored.

Samples of jet events from proton-proton collisions processes are simulated with PYTHIA 8 [14] using a $2 \rightarrow 2$ matrix element at leading order in the strong coupling to model the hard subprocess, *p_T -ordered parton shower to model additional radiation* spectrum, eight samples were generated with different thresholds of the hard-scattering partonic transverse momentum \hat{p}_T . The first six of these samples were mixed taking into account their respective production cross sections.

After the event generation, the events are passed through a full GEANT4 [19] detector simulation with a detailed description of the geometry and the material of the ATLAS experiment, and QGSP_BERT [20, 21] as the set of processes describing the hadronic interactions. The energy deposited by particles in the active detector material is converted into detector signals in the same format as the detector read-out. Finally the Monte Carlo generated events are processed through the trigger simulation package of the experiment, and are reconstructed and analyzed with the same software as for the real data. The simulated data sample used for the analysis (MC11b) gives an accurate description of the pile-up content and detector conditions for the full 2011 data-taking period.

The data samples employed correspond to proton-proton collisions at $\sqrt{s} = 7$ TeV delivered by the LHC and collected by the ATLAS experiment

during 2011. Only data collected during stable beam periods in which all sub-detectors were fully operational are used. After the application of the data quality selection, the total integrated luminosity is about 4.7 fb^{-1} . The LHC performance steadily improved during 2011, surpassing in the second half of the year the design values for several machine parameters. In particular the average number of minimum-bias pile-up events, originating from collisions of additional protons in the same bunch as the signal collision, grew from 3 to 20. This fact will be of importance when discussing the selection of discriminating variables.

For the study of systematic effects and for result comparison, other Monte Carlo samples were utilised. Results were produced with the HERWIG++ generator [7] and with PYTHIA8 using the Perugia tune [22].

5.2 Reconstruction

Both experimental data and simulated events were reconstructed using version 17 of the ATLAS software. In this section we briefly describe the reconstruction of the two key objects used in this analysis, namely jets and tracks.

Jets are reconstructed using the anti- k_t jet algorithm [23] with a distance parameter $R = 0.4$, using calorimeter topological clusters [24] as inputs. Topological clusters are built starting from seed calorimeter cells with a signal at least four times higher than the root-mean-square (RMS) of the total noise contribution, consisting of electronic and pile-up noise corresponding to an average luminosity of $\mu = 8$. Cells neighbouring the seed which have a signal-to-RMS-noise-ratio of two are then iteratively added. Finally, all nearest neighbour cells are added to the cluster without any threshold. Several quality criteria are applied to eliminate jets that are produced by problematic calorimeter behaviour [25]. In particular the jet is rejected if 90% of the jet energy is distributed over less than 6 calorimeter cells (this handles fake jets caused by sporadic noise bursts in the LAr calorimeter), if the fraction of jet energy from LAr calorimeter cells flagged as problematic is greater than 0.8, if the fraction of energy deposited in the electromagnetic calorimeter is greater than 0.95 (this removes fake jets caused by noise bursts) and if the cell-weighted time of the jet is more than 50 ns different from that of the average event time (which can be jets with large out-of-time energy deposits). The jet energies are corrected for inhomogeneities and for the non-compensating nature of the calorimeter by using $p_T - \text{and} \eta$ -dependent calibration factors determined from Monte Carlo simulation [26]. This calibration is referred to as the EM+JES scale. Using test beam results, in-situ

track and calorimeter measurements, estimations of pile-up energy depositions, and detailed Monte Carlo comparisons establish an uncertainty of the absolute jet energy scale smaller than $\pm 10\%$ for $\eta < 2.8$ and $p_T > 20$ GeV. More sophisticated techniques undergoing commissioning, such as local cluster weighting, are expected to considerably improve the jet energy uncertainty and resolution [27].

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 $10\text{ }\mu\text{m}$ in the $r - \phi$ plane and $115\text{ }\mu\text{m}$ along the beam axis (z). Tracks with $p_T^{\text{track}} > 150$ MeV and consistent with the beamspot are associated in primary vertices via a finding/fitting two-stage algorithm. Several primary vertices can be reconstructed per event due to the presence of in-time pile-up. The one with at least five associated tracks, a z position within 100 mm of the ATLAS geometrical center, and the largest $\sum_{\text{trk}} p_T^2$, is selected as the one associated to the hard interaction.

5.3 b -jet Tagging

In this analysis jets are identified as b -quark candidates (b-tagged) by ATLAS MV1 b-tagging algorithm, based on a neural network using the output weights of three advanced taggers: IP3D, SV1 and the combination of JetFitter and IP3D [3]. All these algorithms are based on Monte Carlo predictions for the signal (b-jet) or background (light- or in some cases c-jet) hypotheses. They use the information from the impact parameters of displaced tracks in the jet as well as the topological characteristics of secondary decay vertices reconstructed within it. b -jets were selected using a working point with 60% efficiency for b -quark jets.

5.4 Event and jet selections

All data jets considered in this analysis was collected by the jet trigger chain of the ATLAS 3-level Trigger System. At the hardware Level 1 and local software Level 2, cluster-based jet triggers are used to select events. The last

p_T (GeV)	Run 177531 - 187109	Run > 187109
(40,60)	EF_j20_a4_EFFS OR EF_j15_a4_EFFS	EF_j20_a4tc_EFFS OR EF_j15_a4tc_EFFS
(60,80)	EF_j30_a4_EFFS OR EF_j20_a4_EFFS	EF_j30_a4tc_EFFS OR EF_j20_a4tc_EFFS
(80,110)	EF_j40_a4_EFFS OR EF_j30_a4_EFFS	EF_j40_a4tc_EFFS OR EF_j30_a4tc_EFFS
(110,150)	EF_j55_a4_EFFS OR EF_j40_a4_EFFS	EF_j55_a4tc_EFFS OR EF_j40_a4tc_EFFS
(150,200)	EF_j75_a4_EFFS OR EF_j55_a4_EFFS	EF_j75_a4tc_EFFS OR EF_j55_a4tc_EFFS
(200,270)	EF_j100_a4_EFFS OR EF_j75_a4_EFFS	EF_j100_a4tc_EFFS OR EF_j75_a4tc_EFFS
(270,360)	EF_j135_a4_EFFS OR EF_j100_a4_EFFS	EF_j135_a4tc_EFFS OR EF_j100_a4tc_EFFS
(360,480)	EF_j180_a4_EFFS OR EF_j135_a4_EFFS	EF_j180_a4tc_EFFS OR EF_j135_a4tc_EFFS

stage, the so-called Event Filter, runs the offline anti- k_t jet finding algorithm with $R = 0.4$ on topological clusters over the complete calorimeter. For this analysis, the events are required to come from single-jet triggers with different thresholds at the Event Filter, ranging from 40 GeV to 480 GeV. Except for the highest p_T thresholds, these triggers were increasingly prescaled with the rapid increase in instantaneous luminosity over time. The triggers with the lowest p_T thresholds were prescaled by up to five orders of magnitude, and typically the same jet trigger is prescaled ten times more in the later data taking periods compared to the early ones. A logical OR of these triggers is used for selecting the events. The offline event selection consists of requiring at least one primary vertex candidate to have 5 or more tracks.

All jets, reconstructed with the anti- k_t $R = 0.4$ algorithm built from topological clusters, were required to be in a region with full tracking coverage, $|\eta_{jet}| < 2.1$. Jets were subdivided in eight p_T bins chosen such as to match the ATLAS inclusive jet trigger thresholds of 98% triggering efficiency. The trigger strategy is detailed in Table 5.1. Trigger selection is not applied over Monte Carlo sample events but the same p_T bins were chosen for consistency in comparisons.

Jets were classified as b -jets using the MV1 b -tagging algorithm at the 60% efficiency working point, corresponding to a weight $w > 0.905$. The reconstructed b -tagged jets were further classified into single and merged b -jets based on truth Monte Carlo information. A B hadron is considered to be associated to a jet if the ΔR distance in $\eta - \phi$ space between the direction of the hadron and the jet axis is smaller than 0.4. Jets were labeled as merged (single) b -jets if they contain two (only one) B hadron.

Only b -tagged jets having no close-by jet with p_T higher than 7 GeV at electromagnetic scale were considered for the analysis.

5.5 Track selection

It is important to select genuine tracks belonging to jets. Only tracks located within a cone of radius $\Delta R(jet^{\text{reco}}, \text{track}) \leq R$ around the jet axis were considered. Cuts on $p_{\text{T}}^{\text{trk}} > 1.0 \text{ GeV}$ and $\chi^2/ndf < 3$ are applied as a minimum starting point. In addition, tracks are required to have at least seven precision hits (both pixel or micro-strip) in order to guarantee at least 3 z -measurements. Tracks are also required to fulfill cuts on the transverse plane and longitudinal impact parameters at the perigee to ensure that they arise from the primary vertex. As cutting on impact parameter (IP) significance might be detrimental for b -jets, where large IP values are expected, the relaxed cuts were $|IP_{xy}| < 200 \mu\text{m}$, and $|IP_z \sin \theta| < 200 \mu\text{m}$.

Chapter 6

BB -jet tagging

6.1 Introduction

the strategy. We should explain that the purpose is to find 2 B-hadrons inside a single jet and that there are two different strategies: attempt to reconstruct multiple vertices, or use jet shape/substructure information. You can then discuss the issues with the former approach, in particular the lower efficiency due to the double b-tag requirement that goes as b-tag eff square, and reference CDF. Then say that for this paper we are exploring a different way to tag double B-hadron jets, that does not rely on explicitly finding vertices. Eventually, the goal should be to combine both approaches.

6.2 Kinematic differences between single b - and merged $b\bar{b}$ -jets

The distinct characteristics between b and $b\bar{b}$ jets are expected to arise from the two-prong (two B-hadrons) substructure of merged jets. They are thus expected, for the same jet p_T , to have higher track-multiplicity and be wider than single b -jets. Based on these characteristics, the following calorimeter and tracking variables were studied in simulated QCD samples of b -tagged jets:

- Jet track multiplicity
- Track jet width (p_T weighted)
- Nr. of sub track-jets
- ΔR_{max} (tracks)
- ΔR leading tracks

- ΔR (k_t subjets)
- Jet topological-cluster multiplicity
- Calorimeter width
- Jet Mass
- Eccentricity (track & calo)
- τ_2 : 2-subjettiness
- τ_2/τ_1

Figures 6.1 to 6.10 show the distributions and correlations of some of these variables in selected bins of b -tagged jet p_T .

Figure 6.1 shows the distribution of jet track-multiplicity. This variable is simple to calculate and carries important information of the jet inner structure. As it can be seen merged b -jets contain on average around two more tracks than single b -jets at low jet p_T , with a larger difference at higher p_T values. The track-multiplicity corresponds to tracks with p_T above 1 GeV. The effect of using a minimum track p_T of 0.5 GeV was also examined. This was motivated by the fact that it could lead to an improvement in discrimination if it captured more information about the fragmentation process. On the other hand, a lower minimum track p_T can make the method more sensitive to pile-up with the addition of soft tracks incorrectly associated to the jets. What it was observed is that reducing the p_T cut only widens the distributions without increasing the separation between single and merged jets.

Figure 6.2 shows that, as expected, merged b -jets are wider than single b -jets. The track-jet width was computed as the track p_T weighted average of the ΔR distance between the tracks matched to the jet and the jet axis: $Width = \frac{\sum_{i=1}^N p_T^{trk_i} \Delta R(trk_i, jet)}{\sum_{i=1}^N p_T^{trk_i}}$ where N is the total number of tracks in the jet.

In Fig. 6.3 the correlation between the track-jet width and the track-multiplicity is shown for single and merged b jets. These two variables alone provide a good discrimination for tagging $b\bar{b}$ merged jets.

The calorimeter width, using the topo-cluster jet constituents, gives also good separation. However, this variable is more sensitive to the amount of pile-up in the event than its track-based counterpart. In Fig. 6.4 the distributions of calorimeter width for single and merged b -jets can be seen for events with low and high number of primary vertices (NPV). In Fig. 6.5 the same distributions are shown for the track-jet width. Calorimeter width varies with NPV and due to this behavior the track-based version is more

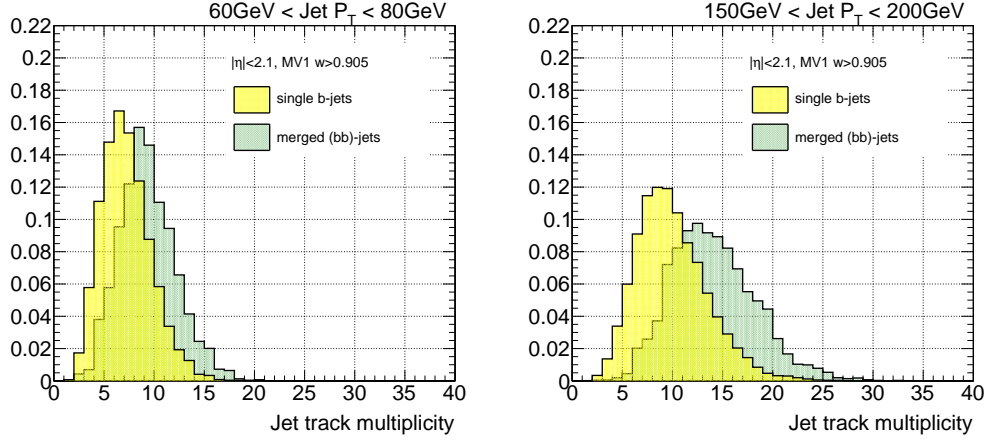


Figure 6.1: Distribution of track multiplicity in jets for single and merged b -jets between 60 GeV to 80 GeV (left) and 150 GeV to 200 GeV (right).

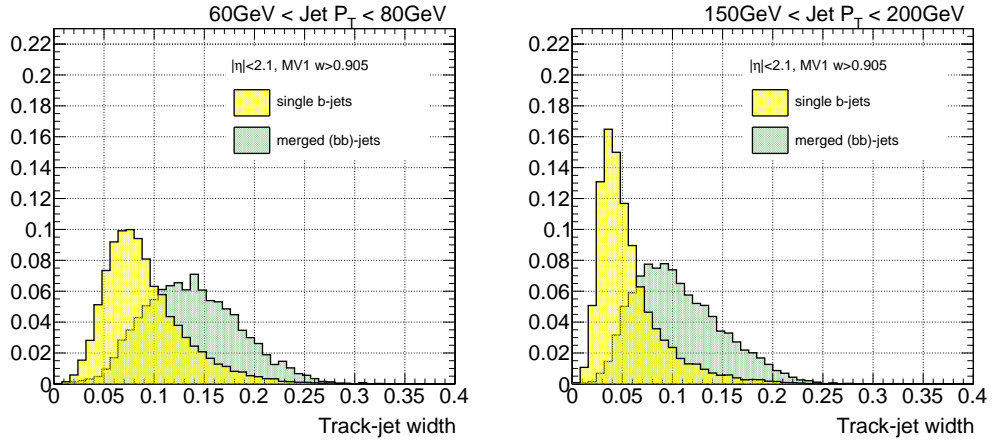


Figure 6.2: Distribution of track-jet width in jets for single and merged b -jets between 60 GeV to 80 GeV (left) and 150 GeV to 200 GeV (right).

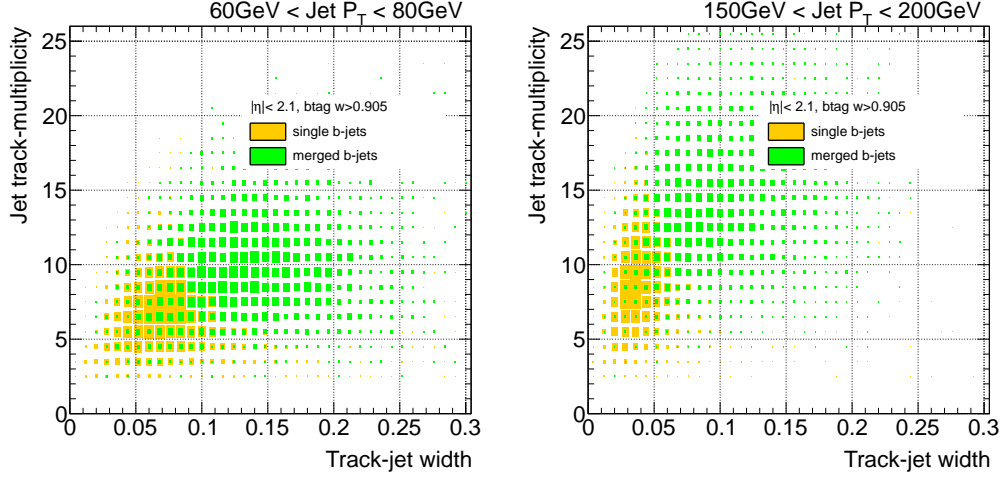


Figure 6.3: Correlation between track-multiplicity and track-jet width for single and merged b -jets between 60 GeV to 80 GeV (left) and 150 GeV to 200 GeV (right).

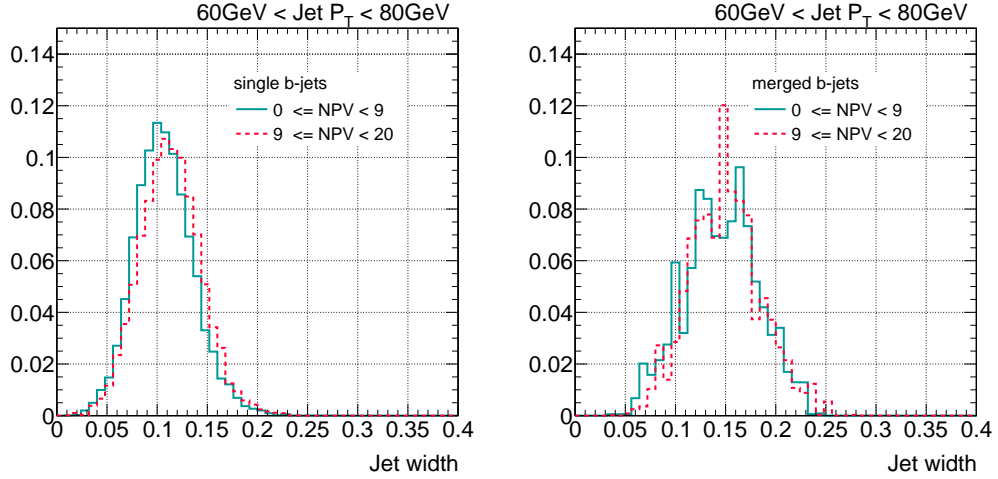


Figure 6.4: Distribution of calorimeter jet width for single (left) and merged (right) b -jets in bins of the Number of Primary Vertices for jets between 60 GeV to 80 GeV.

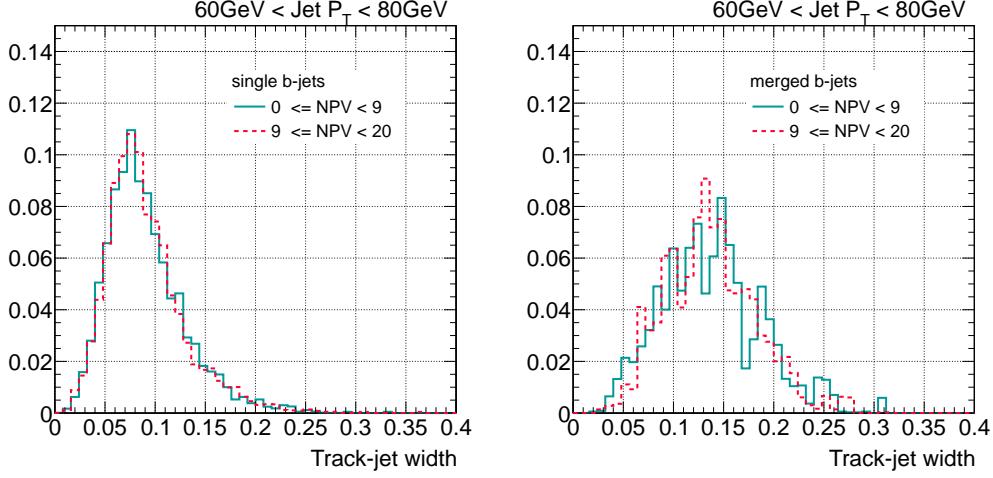


Figure 6.5: Distribution of track-jet width for single (left) and merged (right) b -jets in bins of the Number of Primary Vertices for jets between 60 GeV to 80 GeV.

suitable as a more robust discriminator. For similar reasons, the topo-cluster multiplicity and the jet mass were discarded as discriminating variables.

systematics.ods

Figure 6.6 shows the distribution of the maximum ΔR between track pairs in the jets (ΔR_{\max}). Merged b -jets show significantly higher values for this variable over a broad range of jet p_T . The distinct characteristic of this variable is that the separation between single b -jets and merged does not depend on jet p_T . In spite of its good discrimination power, we have looked for alternatives to ΔR_{\max} as it is not an infrared safe magnitude and is sensitive to soft tracks originating from pile-up.

The distribution of the ΔR between the axis of the two k_t -subjets in the jet is shown in Fig. 6.7 for single and merged b -jets. For this variable, tracks inside the jets are clustered into exactly two subjets using the k_t jet algorithm implemented in fastjet [28]. We observe that it also provides good separation, with the advantage of infrared safeness and insensitivity to pile-up.

N -subjettiness variables, as described in Ref. [29], were originally designed to identify boosted objects, like electroweak bosons and top quarks, decaying into collimated shower of hadrons which a standard jet algorithm would reconstruct as single jets. It is defined as: $\tau_N = \frac{1}{\sum_k p_{T_k} R_0} \sum_k p_{T_k} \min\{\Delta R_{S_1,k}, \Delta R_{S_2,k}, \dots, \Delta R_{S_N,k}\}$ where R_0 is the jet radius used in the jet clustering algorithm and the sum runs over the components of the jet. To avoid dependence on pile-up we consider the track-based subjettiness, where the sum is over the tracks in the b -tagged jet. $\Delta R_{S_j,k}$ is the distance in the rapidity-azimuth plane between

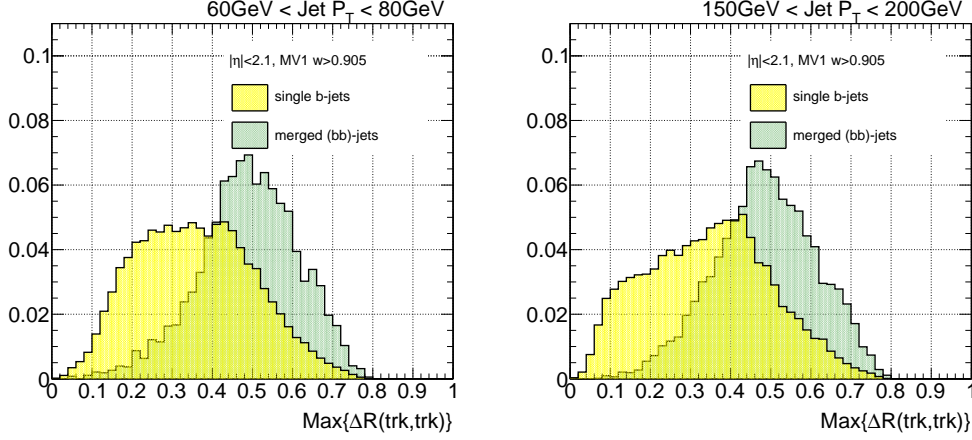


Figure 6.6: Distribution of the maximum ΔR between pairs in jets (ΔR_{\max}) for single and merged b -jets between 60 GeV to 80 GeV (left) and 150 GeV to 200 GeV (right).

the axis of subjet j and constituent track k . This jet shape variable quantifies to what degree a jet can be regarded as composed of N subjets. For instance, a jet with a two pronged structure, with all tracks clustered along two directions, is expected to have a smaller τ_2 value than a jet with tracks uniformly distributed in $\eta - \phi$ space.

Plots of τ_2 are shown in Fig. 6.8. In spite of its expected 2-prong substructure, merged b -jets have higher values of τ_2 than single b -jets. The explanation of this behavior can be found in Fig. 6.9, where its correlation with track-width (τ_1) is shown for single and merged b -jets. The two variables are highly correlated and for this reason wider jets have a larger τ_2 . This suggests to switch from an absolute to a width-normalized τ_2 . Fig. 6.10 thus shows the distributions of τ_2/τ_1 . Although as expected somewhat larger values are obtained for single than for merged b -jets, especially at high p_T , we decided not to use this variable as it offers only marginal discrimination.

Variables such as the ΔR between the two leading tracks in jets and the jet eccentricity did not show good discrimination either and were not considered for the multivariate study.

6.3 Validation of the jet variables in data

In order to study the extent to which the simulation reproduces the distributions observed in data for the different variables explored a set comparison plots is presented. Fig. 6.11 shows the distributions of N_{trk} , track-jet width

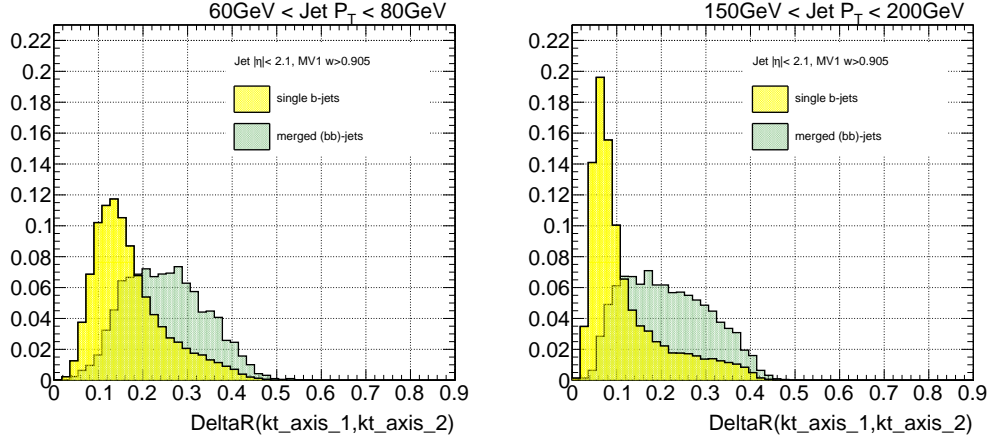


Figure 6.7: Distribution of the maximum ΔR between two k_t axis in jets (ΔR_{\max}) for single and merged b -jets between 60 GeV to 80 GeV (left) and 150 GeV to 200 GeV (right).

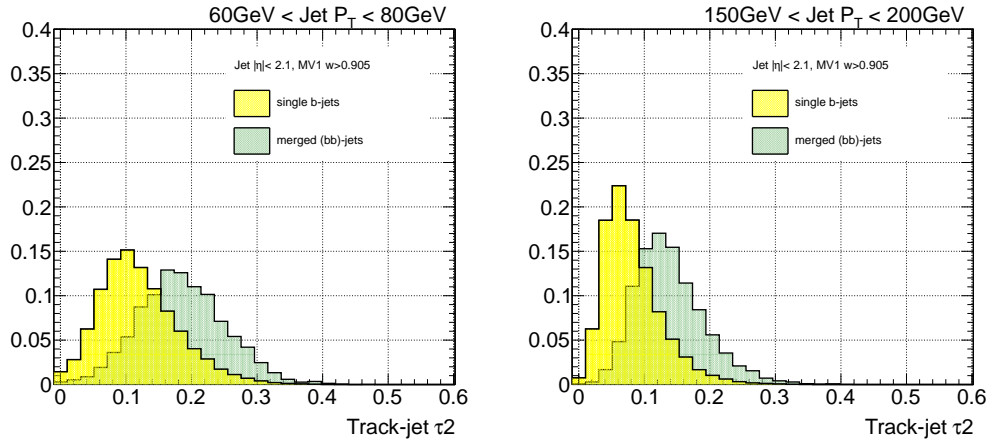


Figure 6.8: Distribution of τ_2 in jets for single and merged b -jets between 60 GeV to 80 GeV (left) and 150 GeV to 200 GeV (right).

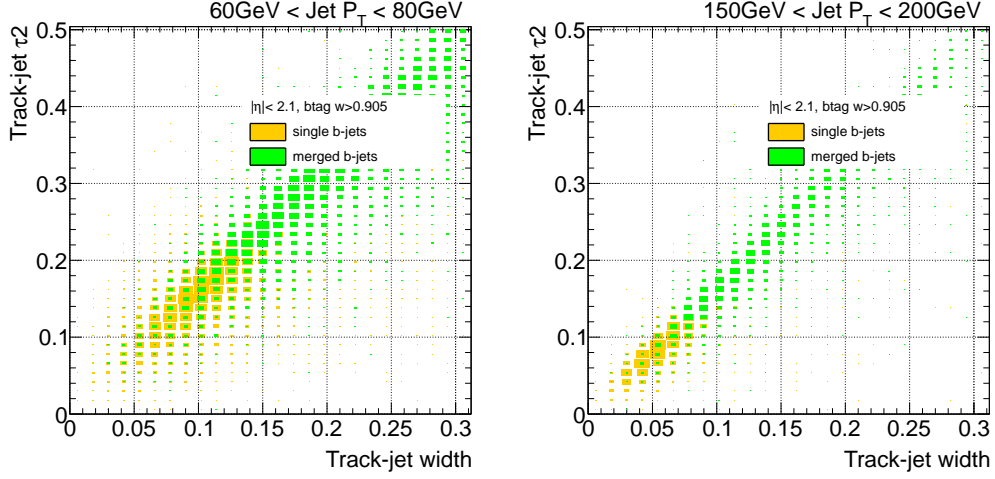


Figure 6.9: Correlation between τ_2 and track-jet width for single and merged b -jets between 60 GeV to 80 GeV (left) and 150 GeV to 200 GeV (right).

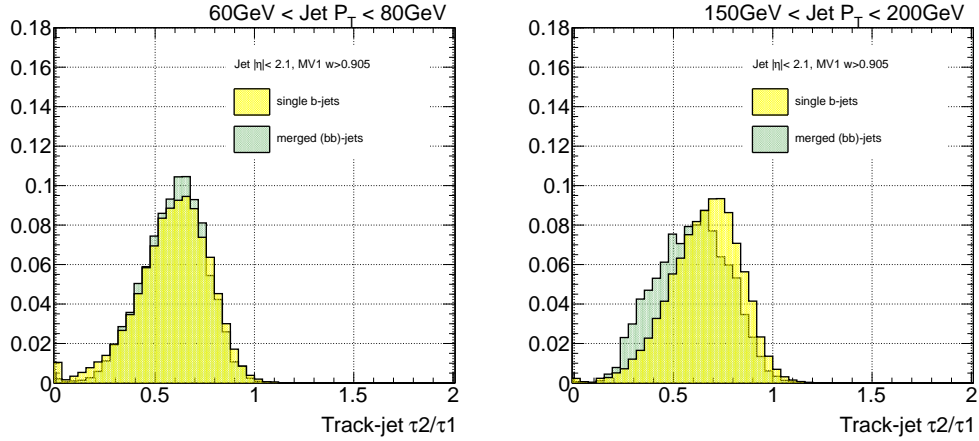


Figure 6.10: Distribution of τ_2/τ_1 in jets for single and merged b -jets between 60 GeV to 80 GeV (left) and 150 GeV to 200 GeV (right).

(p_T weighted) and ΔR_{k_T} in two different jet p_T bins in dijet Monte Carlo and data events from periods B to H overlaid. The distributions are normalized to unit area to allow for shape comparisons. There is a good agreement between data and simulation.

For the sake of evaluating the effect of the increasing amount of pile-up in the last periods of data-taking, the same distributions were built using data events from periods I to M. As expected, no variation was observed. In Fig. 6.12 the track-multiplicity is shown for these periods, in the same two p_T bins for comparison.

6.4 Multivariate Analysis

A discriminant between single b -jets and merged b -jets was built by training a simple likelihood estimator in the context of the Toolkit for Multivariate Data Analysis, TMVA [30].

A sub-set of the di-jet Monte Carlo sample was used for training. After the event and jet selections were performed, the jets with $|\eta| < 2.1$ and $w > 0.905$ were classified as signal (single b -jets) or background (merged b). The likelihood training was done in bins of calorimeter jet p_T . Signal and background jets were not weighted by the dijet samples cross-sections to allow the contribution of subleading lower p_T jets from high p_T events, and thus increase the statistic of merged jets.

Several combinations of the tracking and jet shape variables studied in the previous section were tested as input variables. We found that N_{trk} , track-jet width and ΔR_{k_T} offer the best performance.

The distribution of the likelihood output for single and merged b -jets is shown in Fig. 6.13 for three different bins of calorimeter jet p_T .

The performance of the $g \rightarrow b\bar{b}$ tagger in the simulation can be seen in Fig. 6.14. In this plot the rejection ($1/\epsilon_{bkg}$) of merged b -jets from gluon splitting is shown as a function of single b -jet efficiency for the eight bins of jet p_T mentioned in section 5.4. The performance improves with p_T :

- $p_T > 40$ GeV: over 85% rejection at 50% eff.
- $p_T > 60$ GeV: 90% rejection at 50% eff.
- $p_T > 150$ GeV: over 95% rejection at 50% eff.

The likelihood was trained with jets that had been first tagged by the MV1 algorithm. In order to use the $g \rightarrow b\bar{b}$ classifier for jets tagged by another tagger a new training is required.

The rejection of merged jets attained as a function of p_T for the 50% and 60% efficiency working point distributions. The error is slightly lower for the 60% efficiency working point

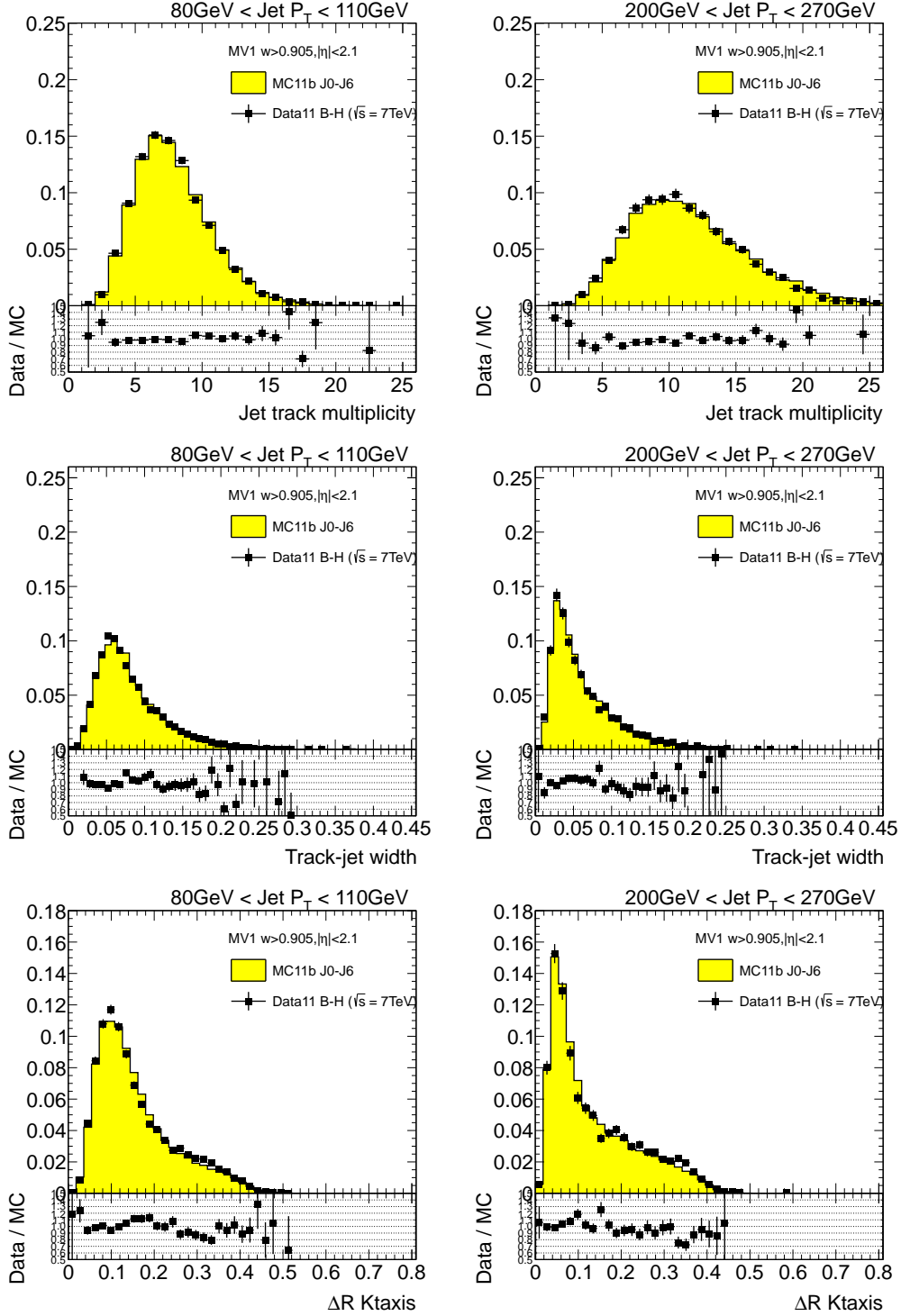


Figure 6.11: Distribution of three tracking variables in 2 different jet p_T bins, for experimental data from data-taking periods B to H (solid black points) and simulated data (filled histograms). The ratio data/simulation is shown at the bottom of the plot.

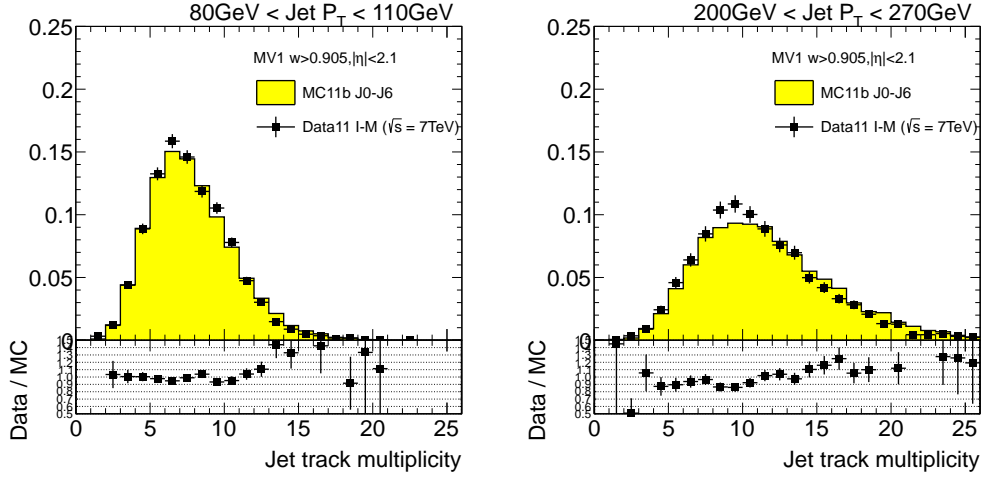


Figure 6.12: Distribution of the jet track-multiplicity in 2 different jet p_T bins, for experimental data from data-taking periods I to M (solid black points) and simulated data (filled histograms). The ratio Data/MC is shown at the bottom of the plot.

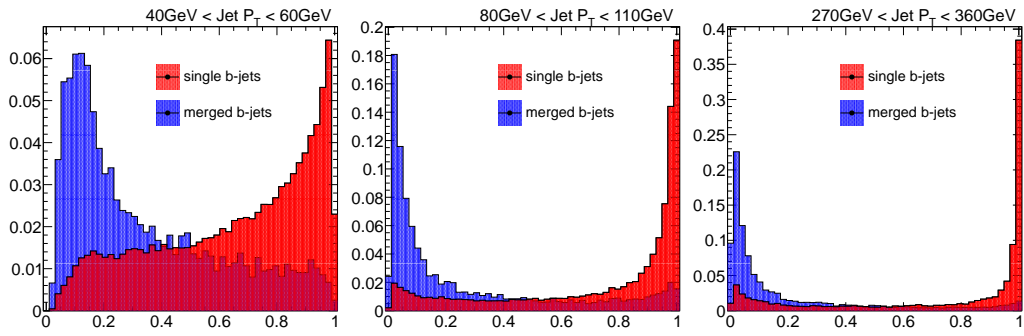


Figure 6.13: Distribution of the $g \rightarrow b\bar{b}$ likelihood output for single and merged b -jets for different jet p_T bins.

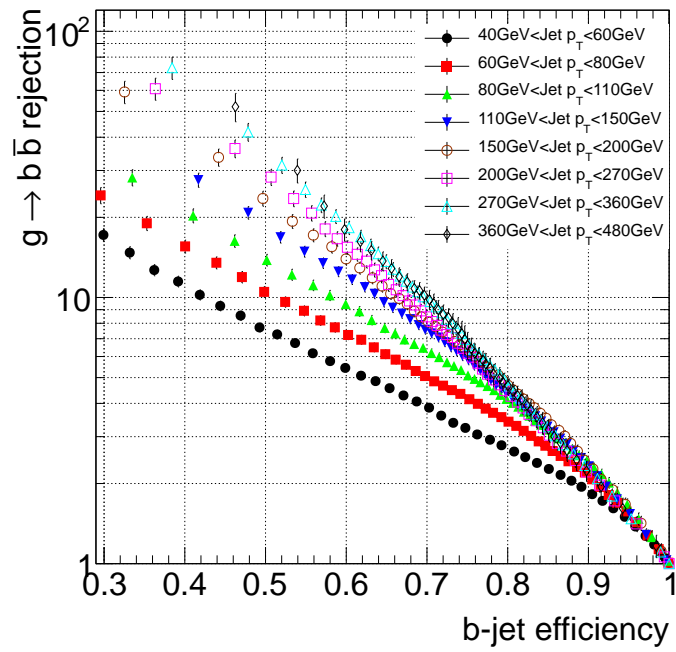


Figure 6.14: Rejection of $g \rightarrow b\bar{b}$ merged b-jets as a function of b-jet efficiency for dijet events in 8 jet p_T bins.

because a higher efficiency allows for a greater number of Monte Carlo events to measure the performance.

6.5 Systematic uncertainties

The development, training and performance determination of the tagger is based on simulated events. Although the agreement between simulation and data explored in section ?? is a necessary validation condition, it is important to investigate how the tagger performance depends on relevant systematics in the data. In particular we have considered:

- presence of additional interactions (pile-up)
- uncertainty in the b -jet tagging efficiency
- uncertainty in the track reconstruction efficiency
- uncertainty in the track transverse momentum resolution
- uncertainty in the jet transverse momentum resolution

I. Pile-up

The size of this effect was studied by comparing the performance of the likelihood discriminant with b -jets in events with small (1-9) and large (9-20) number of primary vertices. As expected from the use of tracking (as opposed to calorimeter) variables no significant dependence with pile-up is observed within statistics. Of the 20 determinations (2 working points with 10 p_T bins each) of performance differences between high and low number of primary vertex events, i

II. b -tagging efficiency

The performance of heavy-flavor tagging in Monte Carlo events is calibrated to experimental data by means of the scale factors (SFs) measured by the b -tagging group. Such a measurement carries a systematic uncertainty, and in order to estimate its effect a conservative approach is followed: the SFs are varied in all the p_T bins simultaneously by one standard deviation both in the up and down directions. The result of this procedure for the distribution of the tracking variables used as discriminant is illustrated in Fig. 6.15.

The effect of the b -tagging calibration uncertainty on the likelihood performance is negligible. This was indeed expected. The scale factors depend on the true flavor of the jet and on its p_T , but these are basically constant in the performance determination bins.

III. Track reconstruction efficiency

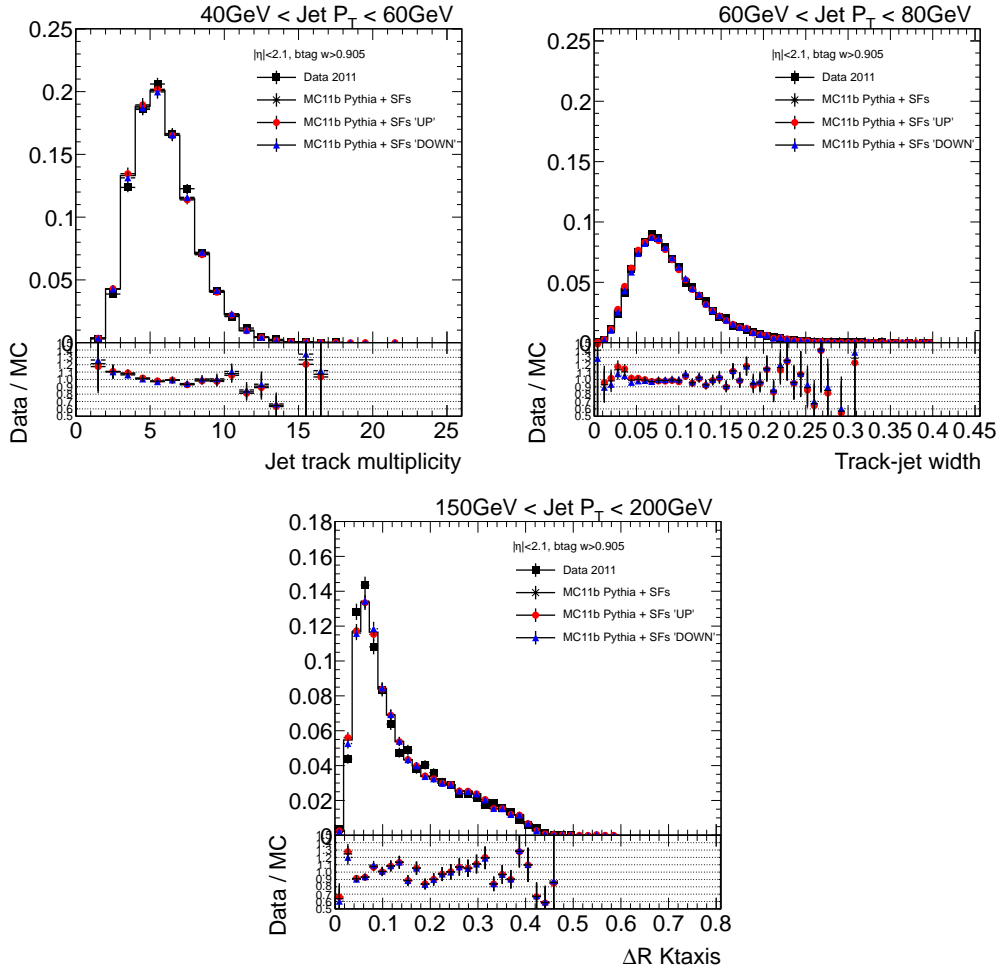


Figure 6.15: The effect of a variation in the b -tagging scale factors on the tracking variables distributions. SFs were varied up (down) by 1-sigma to evaluate the systematic uncertainty from this source.

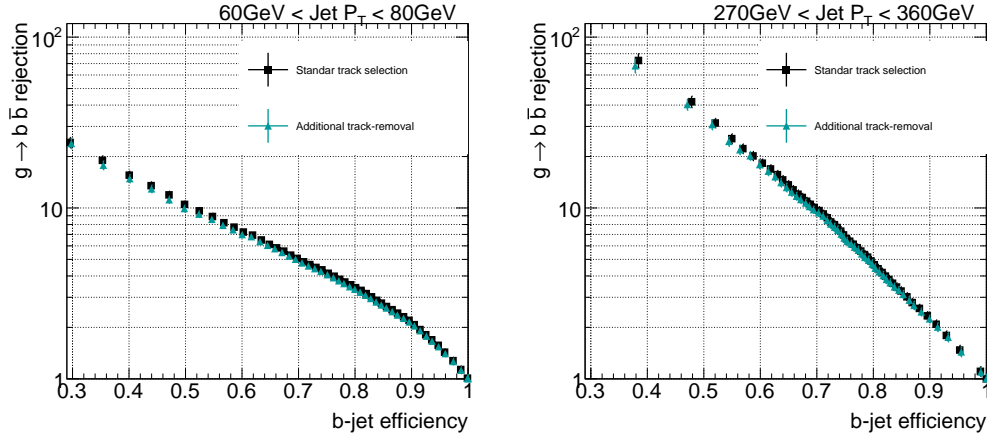


Figure 6.16: Rejection of $g \rightarrow b\bar{b}$ merged b-jets as a function of b-jet efficiency showing shift in likelihood performance caused by rejecting additional tracks.

This uncertainty arises from the limit of in the understanding of the material layout of the Inner Detector. To test its impact a fraction of tracks determined from the track efficiency uncertainty was randomly removed following the method in Ref. [31].

The tracking efficiency systematics are given in bins of track η . For tracks with $p_T^{\text{track}} > 500$ MeV the uncertainties are independent of p_T : 2% for $|\eta^{\text{track}}| < 1.3$, 3% for $1.3 < |\eta^{\text{track}}| < 1.9$, 4% for $1.9 < |\eta^{\text{track}}| < 2.1$, 4% for $2.1 < |\eta^{\text{track}}| < 2.3$ and 7% for $2.3 < |\eta^{\text{track}}| < 2.5$ [32]. All numbers are relative to the corresponding tracking efficiencies.

The tracking variables were re-calculated and the performance of the nominal (pythia) likelihood was evaluated in the new sample with worse tracking efficiency. The rejection-efficiency plots, shown in Fig. 6.16, show a small degradation of the performance which is comparable to the statistical uncertainty. The effect is however systematically present over all 16 p_T bin/workingpoints, without clear p_T dependence. We have thus taken the average over p_T , and obtained

IV. Track momentum resolution

The knowledge of the track momentum resolution is limited by the precision both in the material description of the Inner Detector and in the mapping of the magnetic field. Its uncertainty propagates to the kinematic variables used in the $g \rightarrow b\bar{b}$ tagger. In order to study this effect, track momenta are over-smeared according to the measured resolution uncertainties before computing the rejection. The actual smearing is done in $1/p_T$, with an upper bound to the resolution uncertainty given by $\sigma(1/p_T) = 0.02/p_T$ [33].

The effect is found to be negligible with respect to the statistical uncertainty.

V. Jet transverse momentum resolution

The jet momentum resolution was measured for 2011 data and found to be in agreement with the predictions from the MC11 PYTHIA8-based simulation [34]. The precision of this measurement, determined in p_T and η bins, and, is typically 10%. The systematic uncertainty due to the calorimeter jet p_T resolution was estimated by over-smearing the jet 4-momentum in the simulated data, without changing jet η or ϕ angles. The performance is found to globally decrease by 6%, without a particular p_T dependence.

The different contributions to the systematic uncertainty on the $g \rightarrow b\bar{b}$ rejection are summarized in Table ??.

6.6 Other Monte Carlo generators

The development, training and performance determination of the tagger has been done using Monte Carlo events generated with the PYTHIA8 event simulator, interfaced to the GEANT4 based simulation of the ATLAS detector. An immediate question is what the performance would be if studied with a different simulation. In this section we investigate this question for the Perugia tune of PYTHIA8 and the HERWIG++ event generators.

Fig. 6.17 shows a comparison of the performance-efficiency plots between nominal PYTHIA and the alternative simulations for four selected p_T bins covering the full kinematic range.

The performance in HERWIG shows a systematic trend, with agreement at low p_T and increasingly poor performance compared to pythia as p_T grows. For the Perugia tune, on

The reason for the systematic difference observed between the performances of PYTHIA and HERWIG can be traced to the extent with which jets are accurately modelled. Fig. 6.18 compares the measured jet track multiplicity distributions in b -tagged jets and the prediction from both simulations. It is observed that indeed HERWIG++ does not correctly reproduce the data, particularly at medium and high p_T . Although not shown, the level of agreement is found to be better for jet track width and the angle between the leading subjets, the two other variables used for discrimination.

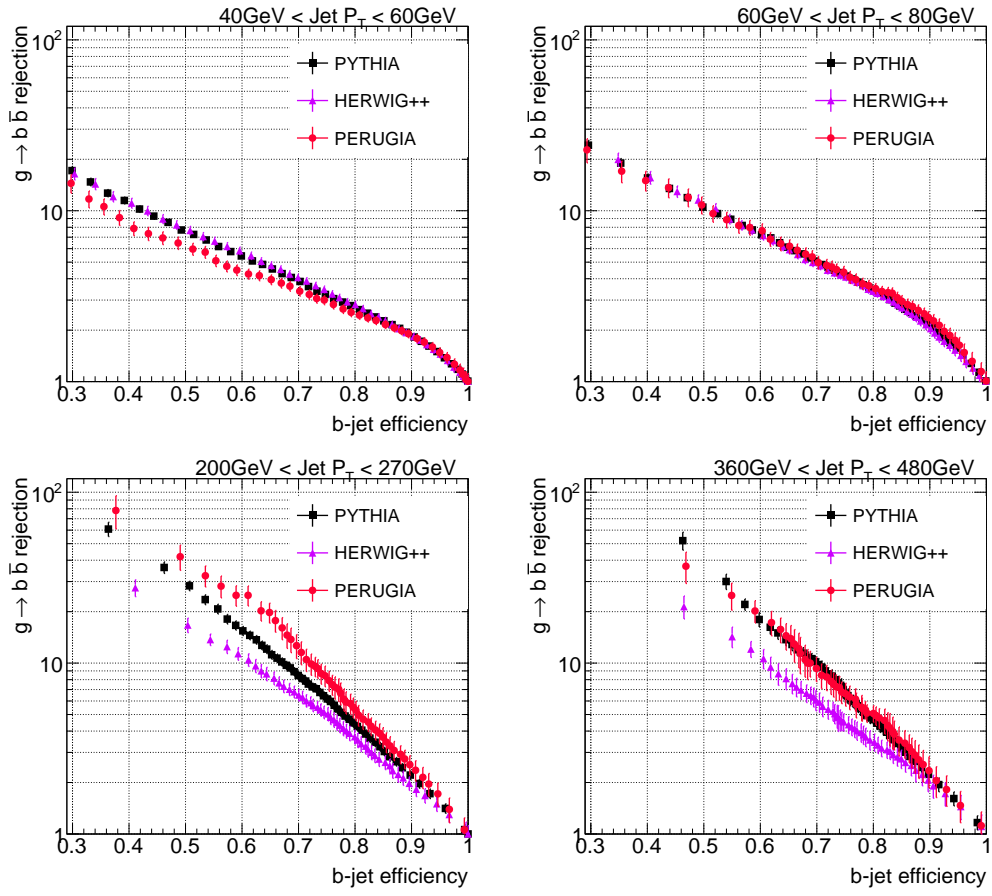


Figure 6.17: Rejection of $g \rightarrow b\bar{b}$ merged b-jets as a function of single b-jet identification efficiency for different Monte Carlo generators.

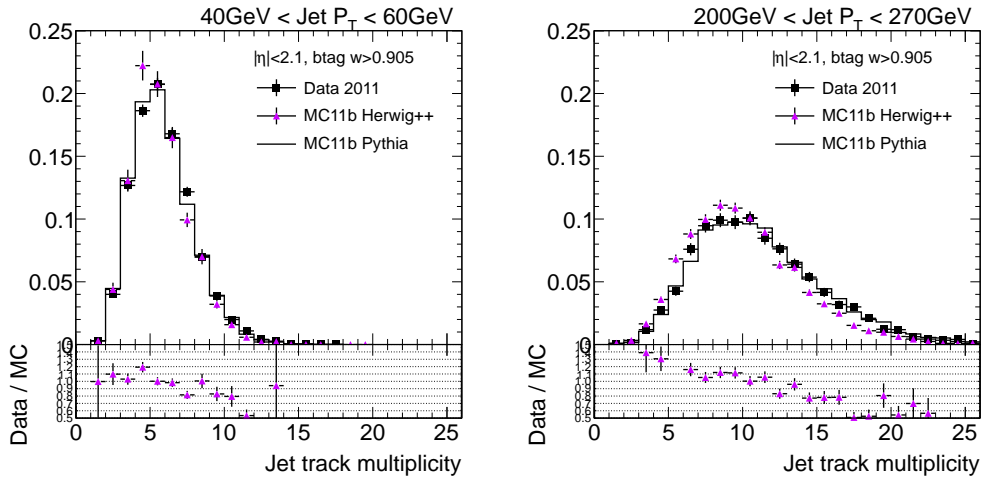


Figure 6.18: Distribution of the jet track-multiplicity in 2 different jet p_T bins, for experimental data from data-taking periods B to H (solid black points) and Herwig++ events (solid violet triangles). The ratio databHerwig++ simulation is shown at the bottom of the plot. Pythia distribution is also shown for reference.

Chapter 7

Fraction of gluon-splitting jets in data

7.1 Template fits

7.2 Systematic uncertainties

Chapter 8

Conclusions

A likelihood classifier has been trained to tag merged b -jets from gluon splitting. The method exploits the kinematic differences between merged and single b -tagged jets using track-based jet shape and jet substructure variables.

This tool provides a handle to understand $b\bar{b}$ production in data and to reduce backgrounds in physics channels involving b -quarks in the final state.

Bibliography

- [1] Performance of impact parameter-based b -tagging algorithms with the atlas detector using proton-proton collisions at $\sqrt{s} = 7$ tev. Technical Report ATLAS-CONF-2010-091, CERN, Geneva, Oct 2010.
- [2] Performance of the atlas secondary vertex b -tagging algorithm in 7 tev collision data. Technical Report ATLAS-CONF-2010-042, CERN, Geneva, Jul 2010.
- [3] Commissioning of the atlas high-performance b -tagging algorithms in the 7 tev collision data. Technical Report ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011.
- [4] Stefano Frixione and Michelangelo L. Mangano. Heavy quark jets in hadronic collisions. *Nucl.Phys.*, B483:321–338, 1997.
- [5] Andrea Banfi, Gavin Salam, and Giulia Zanderighi. Accurate qcd predictions for heavy-quark jets at the tevatron and lh. *JHEP*, 0707:026, 2007.
- [6] S. Catani, Y.L. Dokshitzer, H. Seymour, and B.R. Webber. Longitudinally invariant K(t) clustering algorithms for hadron hadron collisions. *Nucl. Phys.*, B406:187, 1993.
- [7] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001.
- [8] John M. Campbell, R.Keith Ellis, F. Maltoni, and S. Willenbrock. Production of a W boson and two jets with one b^- quark tag. *Phys.Rev.*, D75:054015, 2007.
- [9] Search for supersymmetry in pp collisions at $\sqrt{s} = 7$ tev in final states with missing transverse momentum, b -jets and no leptons with the atlas

- detector. Technical Report ATLAS-CONF-2011-098, CERN, Geneva, Jul 2011.
- [10] A. Abdesselam et al. Boosted objects: a probe of beyond the standard model physics. *The European Physical Journal C - Particles and Fields*, 71:1–19, 2011.
 - [11] A. Altheimer et al. Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks. 2012.
 - [12] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys.Rev.Lett.*, 100:242001, 2008.
 - [13] Atlas detector and physics performance: Technical design report, 1. Technical Report CERN-LHCC-99-14, ATLAS-TDR-14, CERN, Geneva, 1999.
 - [14] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Comput. Phys. Commun.*, 178:852, 2008.
 - [15] R. Corke and T. Sjöstrand. Improved parton showers at large transverse momenta. *European Physical Journal C*, 69:1, 2010.
 - [16] T. Sjöstrand and P. Z. Skands. Transverse-momentum-ordered showers and interleaved multiple interactions. *European Physical Journal C*, 39:129, 2005.
 - [17] B. Andersson et al. Parton fragmentation and string dynamics. *Phys. Rep.*, 97:31, 1983.
 - [18] Atlas tunes of pythia 6 and pythia 8 for mc11. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011.
 - [19] S. Agostinelli et al. Geant4 a simulation toolkit. *Nucl. Inst. Meth. Section A*, 506(3):250 – 303, 2003.
 - [20] G. Folger and J. P. Wellisch. String parton models in Geant4. 2003.
 - [21] H. W. Bertini. Intranuclear-cascade calculation of the secondary nucleon spectra from nucleon-nucleus interactions in the energy range 340 to 2900 MeV and comparisons with experiment. *Phys. Rev.*, 188:1711–1730, 1969.
 - [22] Peter Z. Skands. The Perugia Tunes. 2009.

- [23] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008.
- [24] W Lampl et al. Calorimeter clustering algorithms: Description and performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008.
- [25] Data-quality requirements and event cleaning for jets and missing transverse energy reconstruction with the atlas detector in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 7$ tev. Technical Report ATLAS-CONF-2010-038, CERN, Geneva, Jul 2010.
- [26] Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at $\sqrt{s} = 7$ tev and measured with the atlas detector. Technical Report ATLAS-CONF-2010-056, CERN, Geneva, Jul 2010.
- [27] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. 2009.
- [28] M Cacciari and G.P. Salam. Dispelling the N^3 myth for the k_t jet-finder. *Phys. Lett. B*, 661:057, 2006.
- [29] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N -subjettiness. *JHEP*, 1103:015:026, 2011.
- [30] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. PoS, ACAT:040, 2007.
- [31] R Alon et al. Backup note for measurement of jet mass and substructure in qcd with the atlas experiment. Technical Report ATL-COM-PHYS-2011-401, CERN, Geneva, Apr 2011.
- [32] G. Aad et al. Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC. *New J.Phys.*, 13:053033, 2011.
- [33] Estimating track momentum resolution in minimum bias events using simulation and k_s in $\sqrt{s} = 900$ gev collision data. Technical Report ATLAS-CONF-2010-009, CERN, Geneva, Jun 2010.
- [34] Jet transverse momentum resolution 2011. Technical Report Add this reference if ready before this note., CERN, Geneva, 2012.

McGraw-Hill, 1998.