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

Lic. María Laura González Silva

Tesis Doctoral en Ciencias Físicas

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Noviembre 2012

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

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

Trabajo de Tesis para optar por el título de

Doctor de la Universidad de Buenos Aires en el área Ciencias Físicas

por **María Laura González Silva**

Director de Tesis: Dr. Ricardo Piegaia

Consejero de estudios: Dr. Daniel Deflorian

Lugar de Trabajo: Departamento de Física (CONICET-UBA)

Buenos Aires, 2012

AGRADECIMIENTOS

Quiero agradecer a mi director, Ricardo Piegaia, y a todos aquellos que trabajaron junto conmigo en el experimento ATLAS, Gastón Romeo, Gustavo Otero y Garzón, Hernán Reisin y Sabrina Sacerdotti. Un especial agradecimiento a Ariel Schwartzman y su equipo.

Quiero agradecer también a mis compañeros de grupo y oficina, Javier Tiffenberg, Yann Guardincerri, Pablo Pieroni y Orel Gueta.

Quiero agradecer al Experimento ATLAS, al programa HELEN y al programa e-Planet. Quiero agradecer al CONICET y a la 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 Bejarano y Tomas Teitelbaum.

Quiero agradecer a los amigos que hice a lo largo de estos años en mis visitas al Laboratorio CERN, y a mis colegas y amigos de la Universidad de la Plata. Un especial agradecimiento a Fernando Monticelli.

Quiero agradecer a mis amigos de la vida por continuar a mi lado a pesar de las ausencias.

Finalmente, quiero agradecer a mi familia por su apoyo y comprensión, especialmente a Cristina Silva, Lorena González y Juan Martín Alba.

Abstract

Esta tesis describe un método que permite la identificación de jets que contienen dos hadrones b, que se originan en la división de un gluon en un par $b\bar{b}$. La técnica desarrollada explota las diferencias cinemáticas entre los llamados jets "merged" y los genuinos jets b, usando variables que describen la estructura interna y la forma de los jets, construidas a partir de las trazas asociadas a los mismos. Las variables con mayor poder discriminador son combinadas en un análisis de multivariable. Poder identificar y remover jets b que provienen de la división de un gluon es importante para la estimación y la redución del fondo a señales de física dentro del Modelo Estándar y en nueva física. El algoritmo diseñado rechaza, en eventos simulados, el 95% (50%) de los jets "merged", mientras que retiene el 50% (90%) de los jets b genuinos.

Palabras clave: Experimento ATLAS, Jets, Subestructura de Jets, Etiquetado de Jets b, *Gluon Splitting*.

Abstract

This thesis describes a method that allows the identification of double B-hadron jets originating from gluon-splitting. The technique exploits the kinematic differences between the so called "merged" jets and single B-hadron jets using track-based jet shape and jet substructure variables combined in a multivariate likelihood analysis. The ability to reject b-jets from gluon splitting is important to reduce and to improve the estimation of the b-tag background in Standard Model analyses and in new physics searches involving b-jets in the final state. In the simulation, the algorithm rejects 95% (50%) of merged B-hadron jets while retaining 50% (90%) of the tagged b-jets, although the exact values depend on the jet p_T.

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

Contents

Chapter 1

Double b-hadron jet identification

In this chapter we focus on the understanding of the internal structure of b-jets containing two b-hadrons by investigating the differences between these and single b-quark jets. These differences are expected to arise from the two-subjet structure of double b-hadron or "merged" jets, which would tend to be wider and with a larger number of constituents. Based on these envisaged characteristics, simulated QCD samples of b-tagged jets were used to explore properties with potential discrimination power. The Monte Carlo distributions were compared to data from the 2011 run for validation. We present results from these studies and discuss the choice of the observables selected to build the multivariable tool presented in Chapter ??.

1.1 Data Analysis

The tagging technique presented in this thesis relies on Monte Carlo predictions for the signal (single b) or background (merged b) hypotheses. The accuracy of the simulation is validated with data by comparing the distributions of the different variables studied.

The data samples employed correspond to proton-proton collisions at $\sqrt{s} = 7$ TeV delivered by the LHC and recorded by ATLAS between May and November 2011, with the LHC running with 50 ns bunch spacing, and bunches organized in bunch trains. 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 surviving data corresponds to an integrated luminosity of 4.7 fb^{-1}. The LHC instantaneous luminosity steadily increased during 2011. As a result, 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 from 3 to 20. This fact will be of importance when discussing the selection of discriminating variables.

The Monte Carlo event generators discused in Section **??** are used here. Samples of dijet events from proton-proton collision processes were simulated with PYTHIA version 6.423 [1], used both for the simulation of the hard $2 \to 2$ process as well as for the parton shower, underlying event, and hadronization models. The ATLAS AMBT2 tune of the soft model parameters was used [2]. In order to have sufficient statistics over the entire p_T spectrum, eight samples were generated with different thresholds of the hard-scattering partonic transverse momentum $\hat{p}_T$. Events from different samples were mixed taking into account their respective production cross sections. The simulated data sample used for the analysis gives an accurate description of the pile-up content and detector conditions for the full 2011 data-taking period.

1.1.1 Event selection

The event selection and quality criterion used to extract, from the data and Monte Carlo samples, the final set of jets for the analysis comprises different steps:

- **Trigger**. The event sample was collected using the ATLAS single jet triggers which select events with at least one jet with transverse energy above a given threshold. At the hardware Level 1 and local software Level 2 (see Section **??**), cluster-based jet triggers are used to select events with high-*pt* jets. The Event Filter, in turn, runs the offline anti-k_t jet finding algorithm with $R = 0.4$ on topological clusters over the complete calorimeter. At this stage, the transverse energy thresholds, expressed in GeV, are: 20, 30, 40, 55, 75, 100, 135, 180. These triggers reach an efficiency of 99% for events having the leading jet with an offline energy higher than the corresponding trigger thresholds by a factor ranging between 1.5 and 2. The jet triggers with the lowest p_T thresholds were prescaled by up to five orders of magnitude.

- **Primary vertex**. The offline event selection requires at least one primary vertex candidate with 5 or more tracks. No requirements are placed on the longitudinal position (along the beam line) of the vertex as the beam spot is used as a constraint when fitting the vertex.

- **Primary jet algorithm**. The jet algorithm selected for the analysis was the ATLAS default anti-k_t algorithm [3], with a distance parameter $R = 0.4$, using calorimeter topological clusters [4] as input.

- **Jet calibration**. The EM+JES calibration scheme, described in Section**??**, was used to correct the jet energies for inhomogeneities and for the non-compensating nature of the calorimeter.

- **Jet quality**. Several quality criteria are applied to eliminate "fake" jets that are caused by noise bursts in the calorimeters and energy depositions belonging to a previous bunch crossing [5].

- **Jet tagging**. Only jets tagged as b-jets using the MV1 b-tagging algorithm at the 60% efficiency working point were considered.

- **Isolation**. b-tagged jets with close-by jets ($\Delta R < 0.8$) with p_T higher than 7 GeV at electromagnetic scale were not included in the analysis.

All jets, with transverse momentum between 40 and 480 GeV, the selected p_T range for the analysis, were required to be in a region with full tracking coverage, $|\eta_{jet}| < 2.1$, and they were classified in eight p_T bins chosen such as to match the jet trigger 99% efficiency thresholds (in GeV): 40, 60, 80, 110, 150, 200, 270, 360. An event is used if it satisfies the highest threshold trigger that is 99% efficient for the p_T bin that corresponds to the p_T of its leading jet.

In the case of MC, 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:

$$\text{single } b\text{-jets: } \Delta R(j, b_{1/2}) < 0.4 \tag{1.1}$$

$$\text{merged } b\text{-jets: } \Delta R(j, b_1) < 0.4 \,\&\, \Delta R(j, b_2) < 0.4 \tag{1.2}$$

where j is a jet in the event and $b_{1/2}$ are the b-hadrons in the event. In the case another size parameter is used for jet finding, the definitions in equations 1.1 and 1.2 change accordingly.

1.1.2 Track selection

It is important to select genuine tracks belonging to jets. Only tracks located within a cone of radius $\Delta R(j, \text{track}) \leq 0.4$ around the jet axis were considered. Cuts on ${p_T}^{\text{trk}} > 1.0$ GeV and the χ^2 of the track fit, $\chi^2/ndf < 3$, are applied. In addition, tracks are required to have a total of at least seven precision hits (pixel or micro-strip) in order to guarantee at least 3 z-measurements. Tracks are also required to fulfill cuts on the transverse 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, relaxed cuts were used, $|d_0| < 2$ mm, and $|z_o \sin\theta| < 2$ mm, with θ being the polar angle measured with respect to the beam axis. The track quality cuts are summarized in table 1.1.

Track parameter	Selection
p_T	> 1 GeV
d_0^{PV}	< 2 mm
$z_0^{PV} \sin\theta$	< 2 mm
$\chi^2/ndof$	< 3
Number of Pixel hits	≥ 2
Number of SCT hits	≥ 4
Number of Pixel+SCT hits	≥ 7

Table 1.1: Track selection criteria used for double b-hadron jet tagging, where d_0^{PV} and z_0^{PV} denote the transverse and longitudinal impact parameters derived with respect to the primary vertex. The $chi^2/ndof$ is that of the track fit.

1.2 Kinematic differences between single and double b-hadron jets

The differences between genuine b-quark jets and double b-hadron jets, that in QCD originate mainly from gluon splitting, are expected to arise from the two-subjet structure of merged jets. In this section we present the study of a set of jet shape and substructure variables for the discrimination between single and merged b-jets. These variables are built from jet constituents either at calorimeter level (topological clusters) or tracks associated to the jet.

Jet track multiplicity The jet track multiplicity is a variable simple to calculate that carries important information of the jet inner structure. It is defined as the number of tracks with p_T above 1 GeV, satisfying the quality cuts described in section 1.1.2, and contained within a cone of radius $R = 0.4$ around the jet axis. Figure 1.1 shows its distribution for two p_T bins, representative of the range covered in this study. It is observed that 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 effect of the minimum track p_T requirement was examined by lowering the selection cut to $p_T > 0.5$ GeV. On the one hand this 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. It was observed that reducing the p_T cut of the tracks degrades the discrimination because it widens the distributions without increasing the separation between single and merged jets.

In order to evaluate the effect of using displaced tracks originating from

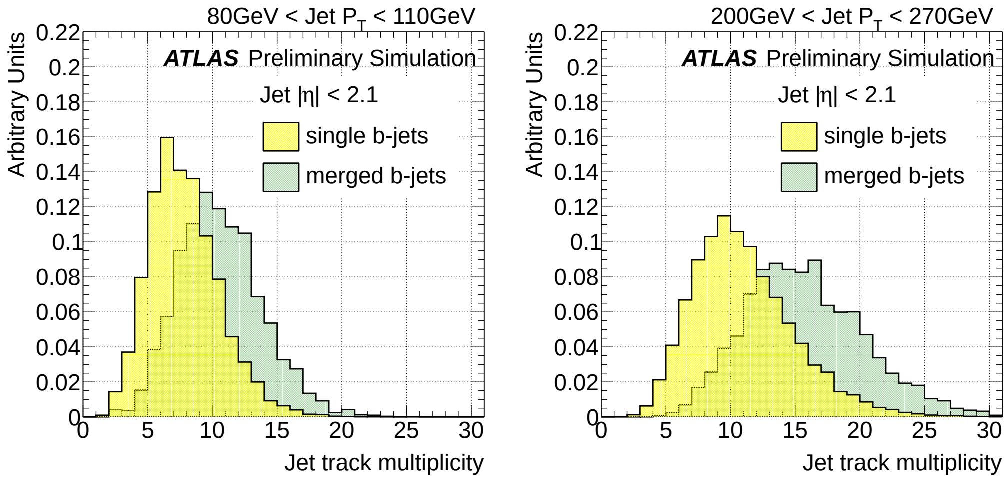
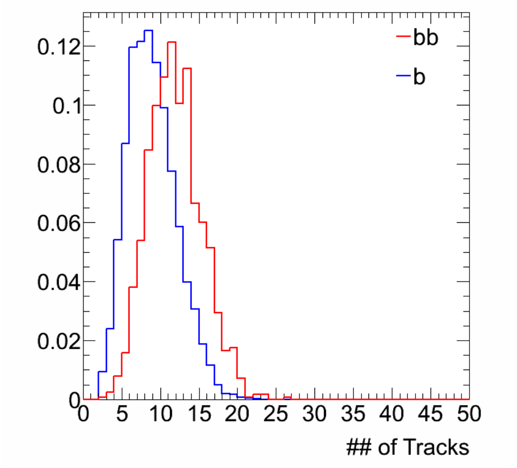


Figure 1.1: Distribution of the track multiplicity in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

the b-hadrons decays a preliminary study was done using a sample of dijet events generated with PYTHIA and with no detector simulation. For this study jets were reconstructed using all stable particles in the event, clustered with the anti-k_t algorithm. The association of charged particles was done in the same way as with the full ATLAS simluation. Distributions of the track multiplicity built using all charged particles and using only charged particles coming from the b-hadron decay ("b-tracks") are illustrated in in Fig. 1.2. A better discrimination between single and merged b-jets is observed when using b-tracks only. The result obtained with PYTHIA suggests that a potential improvement in single-merged separation can be achieved by circumscribing the track selection, in the full simluation, to tracks with large impact parameter significance. A comparison of track multiplicity distributions using all tracks and distributions built with displaced tracks is shown in Fig. 1.3. No improvement is obtained by using displaced tracks ($|d_0|/\sigma(d_0) > 2.5$). The potential sensitivity achieved by enriching the sample in tracks associated to



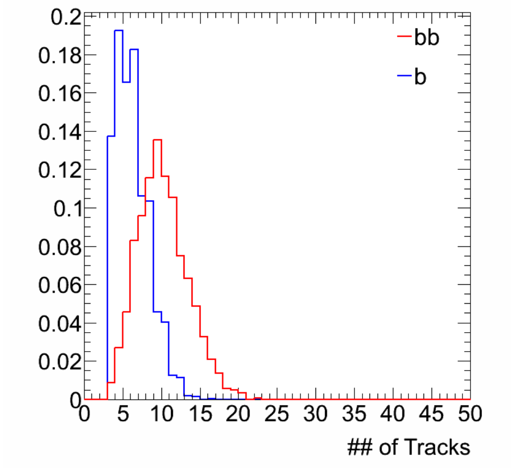


Figure 1.2: Distribution of the track multiplicity in jets for single (b) and merged (bb) jets between 80 GeV to 120 GeV in a sample of dijet events generated with PYTHIA and no detector simulation. Distributions are shown using all charged particles (left) and using only charged particles coming from b-hadron decay (right).

the b-hadron is counterbalanced by the lower number of associated tracks.

Jet width The jet width is part of a set of continuous variables that try to distinguish individual particles/subjets within the jet by means of smooth funcion of $(\Delta\eta, \Delta\phi)$ away from the jet axis, in order to form combinations like geometric moments. This particular combination is a linear moment which sums the distances between the jet constituents and its axes, weighted by the constituent p_T, and then normalized to the total p_T of the jet. Its definition is,

$$Jet\ width = \frac{\sum_{i=1}^{N} {p_T}^{const_i}\, \Delta R(const_i, jet)}{\sum_{i=1}^{N} {p_T}^{const_i}} \tag{1.3}$$

where N is the total number of calorimeter or track constituents.

This observable has been extensively used to discriminate between gluon initiated and light quark initiated jets, see for instance [6] and [7]. Gluon jets

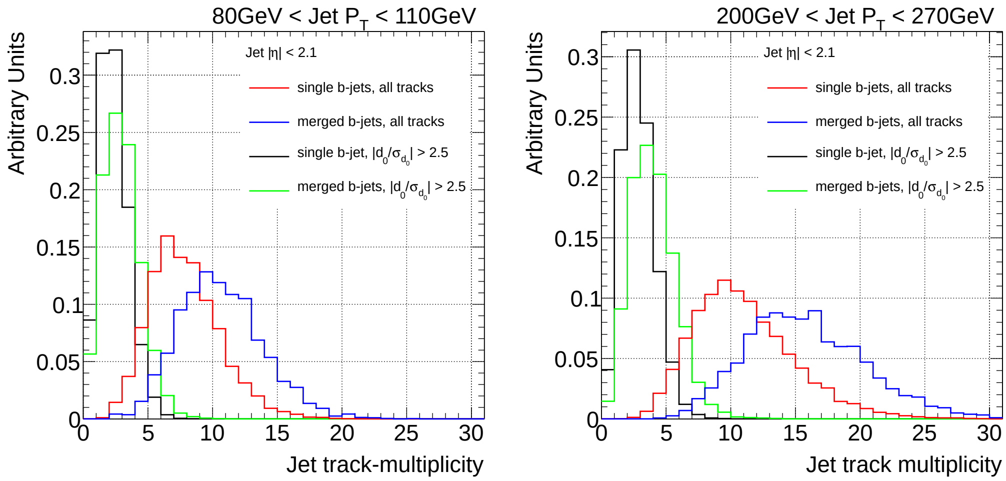


Figure 1.3: Distribution of the jet track multiplicity single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right), for all and displaced tracks only.

are seen to be broader than quark jets. In the case of jets originating from b-quarks, these resemble gluon jets more closely than quarks jets [8]: due to the longer decay chain of b-hadrons the angular spread is larger for a b-jet than a light-quark jet. In order to explore how merged jets, originating from a gluon splitting into a $b\bar{b}$ pair, compare to single b-quark jets and pure gluon jets, a standalone PYTHIA analysis was performed. Figure 1.4 illustrates the result; b-jets containing two b-hadrons present a greater angular width relative to single b-jets and gluon initiated jets. The latter, in turn, look broader than single for medium p_T jets. This behavior is somehow expected in the LHC's higher p_T jets because the QCD shower produces more particles, resulting in broader gluon jets, whereas the particle multiplicity is relatively fixed in the b-hadron decay.

Distributions for the track-jet width with full ATLAS simulation are shown in Fig. 1.5. In this case the sum in equation 1.3 runs over the N

tracks associated to the jet, using the same criteria as for the jet track multiplicity. As expected, merged b-jets are wider than single b-jets.

Pythia standalone samples were also to used evaluate the potential gain in discrimination obtained by utilising all stable particles in the event to build the observable, as opposed to using the charged particles only. A 10% improvement in merged b-jet rejection (for a 50% efficiency in selecting single b-jets) was achieved

The jet width can be measured in terms of calorimeter variables in the full simulation, replacing tracks by topological clusters in the sum. Although it offers good separation, this variable is more sensitive to the amount of pile-up in the event than its track-based counterpart. This is illustrated in Fig. 1.6, which shows the distribution of calorimeter width and track-jet width for single b-jets in events with low and high number of primary vertices (NPV) in a low p_T region where the effect of pile-up is more important.

In general, all the studied calorimeter-based jet variables show similar dependences with NPV. For this reason the track-based versions are preferred as more robust discriminators.

Jet Mass The jet mass, like the linear radial moment, also depends on the radiation pattern of the event. It is the most basic observable for disinguishing massive boosted objects from jets originating from quarks or gluons. The latter are expected to be dominated by wide-angle emissions, with increase probability to see high mass jets initiated from gluons as opposed to quarks [9]. Figure 1.7 shows the distribution of the jet mass for single and merged b-jets. Merged jets tend to have higher masses than single b-jets for the same p_T bin. Although it shows good separation, this calorimeter based variable is susceptible to the amount of pile-up in the event and for this it is not a robust discriminator.

FIGS/VarsSingleMerged/trkWidth_bb_b_g_J3_PT80.png

Figure 1.4: Distribution of track-jet width in jets for gluon-initiated (g), single (b) and merged (bb) jets between 80 GeV to 120 GeV in a sample of dijet events generated with Pythia and no detector simulation.

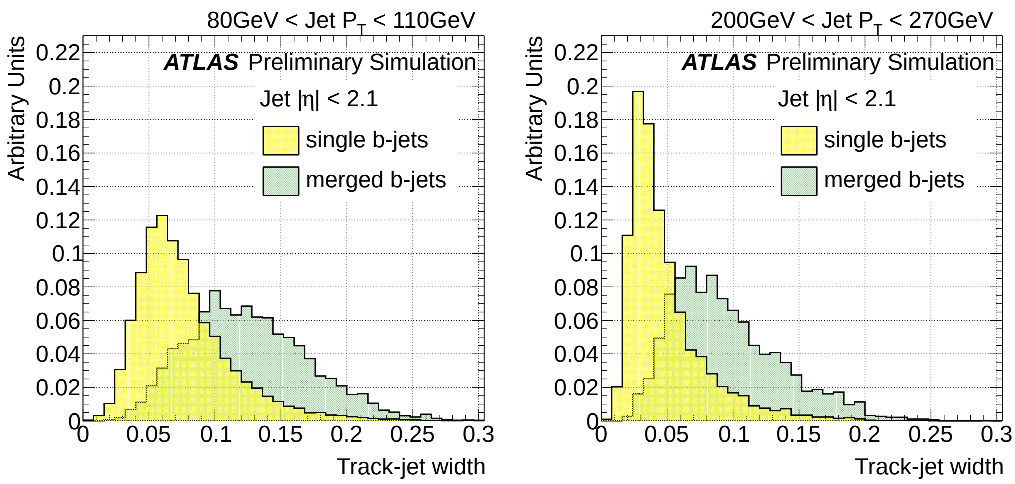


Figure 1.5: Distribution of track-jet width in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

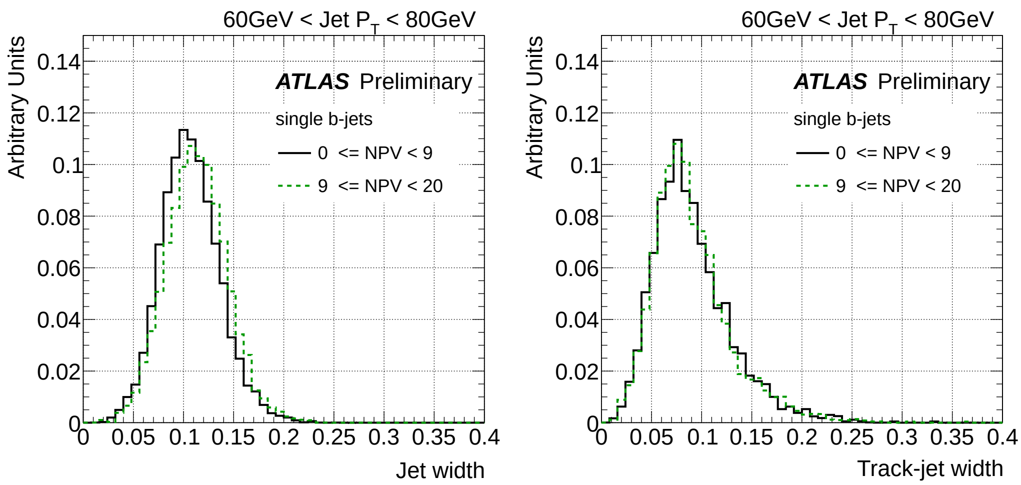


Figure 1.6: Distribution of jet width using topological clusters (left) and tracks (right) for single b-jets in two bins of number of primary vertices (NPV) for jets between 60 GeV to 80 GeV.

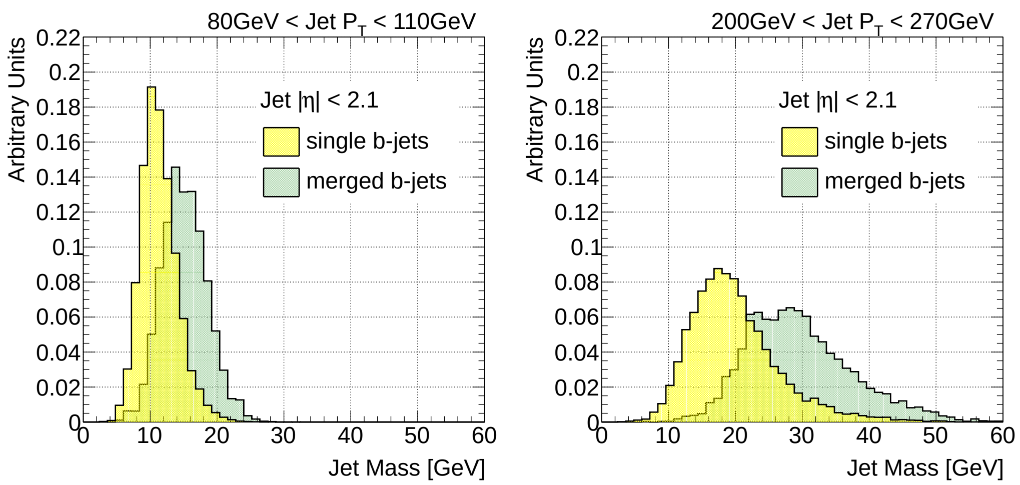


Figure 1.7: Distribution of jet mass in GeV for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

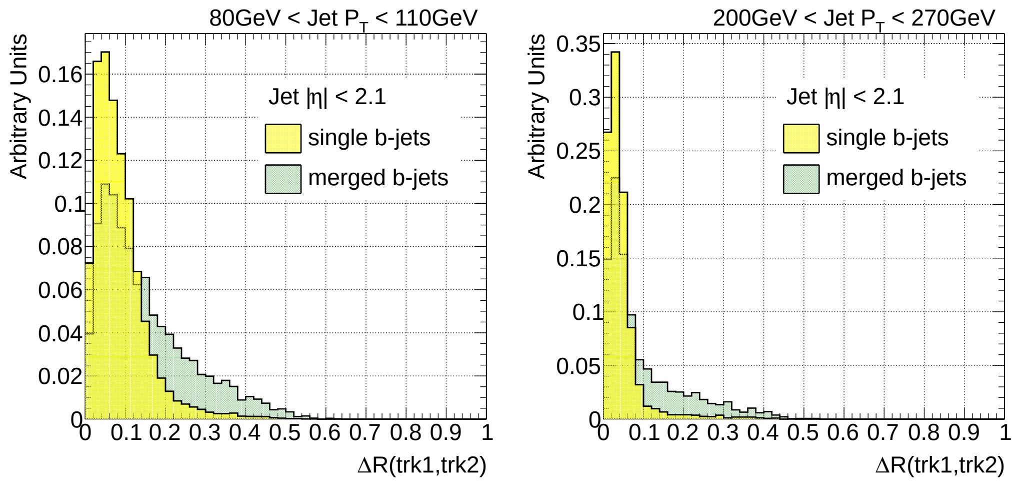


Figure 1.8: Distribution of ΔR between leading tracks for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

ΔR between leading tracks An alternative approach to measuring the width is to use the angular separation of the two hardest constituents inside jets. This has the advantage of removing any dependence on the shower development within the calorimeter and focuses on the hard components of the jet. Figure 1.8 shows the distribution of the ΔR between leading tracks in the jet for single and merged b-jets. The merged b-jet distributions are slightly broader than single b-jet distributions for medium p_T. The effect diminishes as we go to higher transverse momentum values, offering very poor discrimination.

Maximum ΔR between track pairs Several other variables, besides the jet width, were investigated to expose the expected two-subjet substructure of merged b-jets. The maximum ΔR separation between pairs of tracks associated to the jet ($\max\{\Delta R(trk, trk)\}$) was also evaluated as a discrimanting variable. Its distribution is shown in Fig. 1.9, for single and double b-hadron jets. The latter shows significantly higher values for this variable over a broad

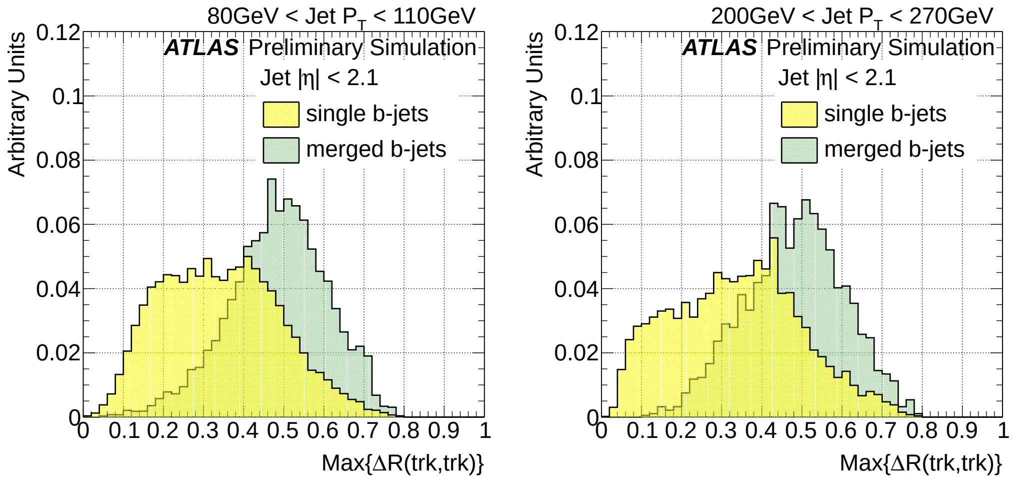


Figure 1.9: Distribution of the maximum ΔR between pairs of tracks in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

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, alternative characterising variables are desirable as $\max\{\Delta R(trk, trk)\}$ is sensitive to soft tracks originating from pile-up.

Number of k_t subjets The subjet multiplicity – the number of subjets within a jet – provides information on the distribution of energy and multiplicity of particles within a jet. For instance, in [10] the result of meassuring this variable on quark- and gluon-initiated jets indicates that gluon-initiated jets tend to have on average higher subjet multiplicity. This result is consistent with the QCD prediction that gluons radiate more than quarks. In the case of this and different other analyses the k_t algorithm is rerun for subjet finding.

By using the sequential recombination algorithms introduced in Section **??**,

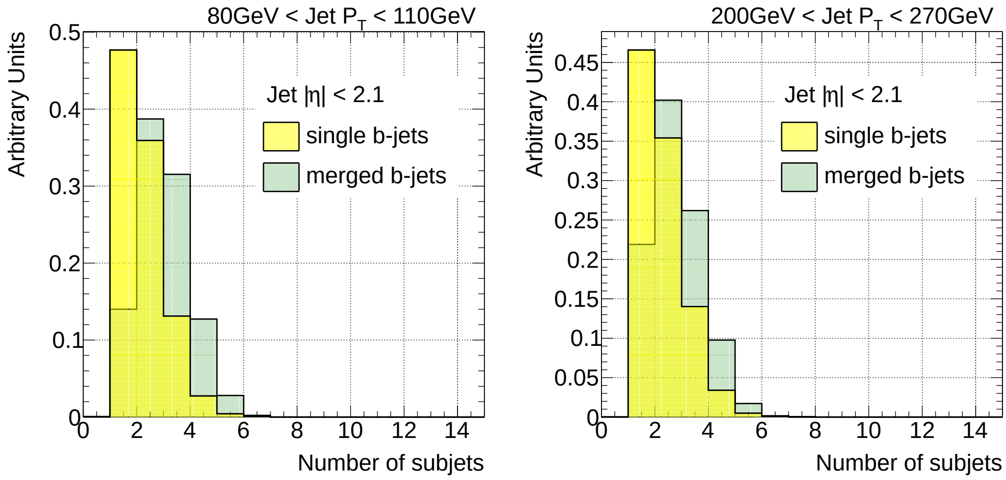


Figure 1.10: Distribution of the number of k_t sub-track-jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

it is straightforward to define a "subjet algorithm" in which the structure of the jet's constituents is resolved using either the same jet finder algorithm as used for jet reconstruction or a new one with a fixed (smaller) distance parameter. As an alternative to fixed distance parameter subjets, it is also possible to undo the last step in the recombination sequence [11] in order to identify the decay products of an object. This approach is used in seveal jet grooming procedures[1], see for instance [13].

Figure 1.10 shows the distribution of the number of subjets for single and merged b-jets. The subjets in this case were built using the associated tracks as constituents, clustered by the inclusive k_t algorithm with distance parameter $R = 0.2$. The discrimination power of this variable, as it is defined, turned out to be very poor.

[1] Jet grooming comprises dedicated techniques to remove uncorrelated radiation within a jet. A review of these procedures can be found in [12].

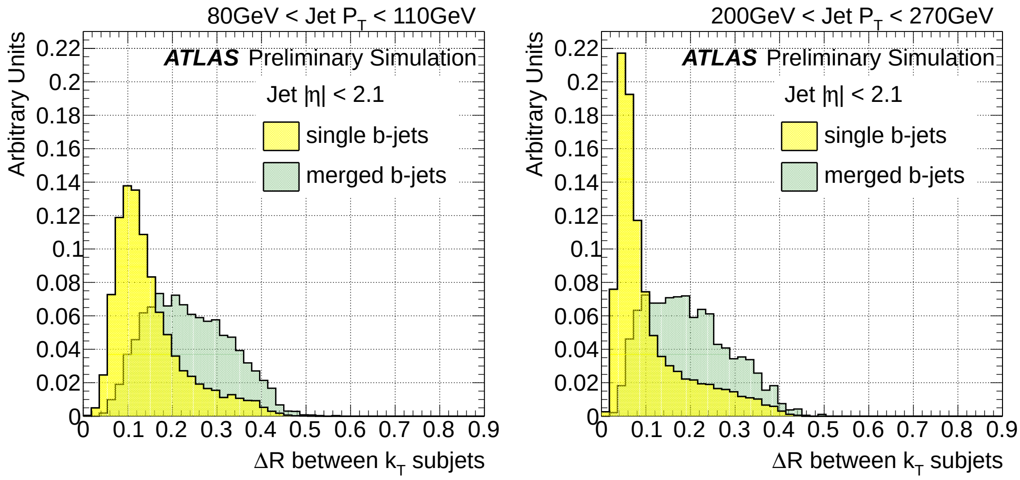


Figure 1.11: Distribution of the ΔR between the axes of the two k_t subjets in the jet for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

ΔR between the axes of two k_t subjets The ΔR between k_t subjets is obtained by applying the exclusive k_t algorithm [14] to the tracks associated to the jet using a large k_t distance parameter to ensure that all tracks get combined. The clustering is stopped once it reaches exactly two jets. The ΔR between the axes of the two exclusive subjets is shown in Fig. 1.11. As expected, it is larger for merged than for single jets. We observe that this variable provides very good separation, with the advantage of infrared safety and insensitivity to pile-up as opposed to $\max\{\Delta R((trk, trk)\}$.

N-subjettiness variables It is possible to extend the use of individual subjets in conjunction with more traditional jet shape variables. Using these tools, an inclusive jet shape based on the substructure topology of a single jet, "N-subjettiness" [15] is defined. This variable describes the energy flow within a jet, quantifying the degree to which radiation is aligned along specified subjet axes. This jet shape was adapted from the event shape N-

jettiness [16].

Given candidate subjets directions determined by an external algorithm such as the exclusive k_t procedure, the variable is defined as,

$$\tau_N^{(\beta)} = \frac{1}{\sum_k p_{Tk}\,(R_0)^\beta} \sum_k p_{Tk} (\min\{\Delta R_{j1,k},\, \Delta R_{j2,k}, ...,\, \Delta R_{jN,k}\})^\beta. \qquad (1.4)$$

The sum runs over the k constituent particles in a given jet where $p_{T,k}$ are their transverse momenta, and $\Delta R_{j1,k}$ is the distance between the candidate subjet $j1$ and a constituent particle k. R_0 is the characteristic jet radius used in the original jet clustering algorithm. The exponential weight, β, can optionally be applied to the angular distance computed between the subjets and the jet constituents. Since eq. 1.4 is linear in each of the constituent particle momenta, this variable is an infrared- and colliner-safe observable.

This jet shape was designed to separate boosted hadronic objects, like electroweak bosons and top quarks decaying into collimated showers of hadrons which a standard jet algorithm would reconstruct as single jets. A simple cut on the ratio τ_N/τ_{N-1} provides excellent discrimination power for N-prong hadronic objects[15] . In particular, τ_2/τ_1 can identify boosted W/Z and Higgs bosons, with the angular weighting exponent $\beta = 1$ providing the best discrimination.

The definition of N-subjettiness is not unique, and different choises can be used to give different weights to the emissions within a jet. The initial step of choosing candidate subjet axes is in fact unnecessary; the quantity in equation 1.4 can be minimised over the candidate subjet directions, further improving boosted object discrimination.

To avoid dependence on pile-up we consider track-based n-subjettiness, where the sum is over the tracks in the b-tagged jet. As seen for massive boosted objects, 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 a

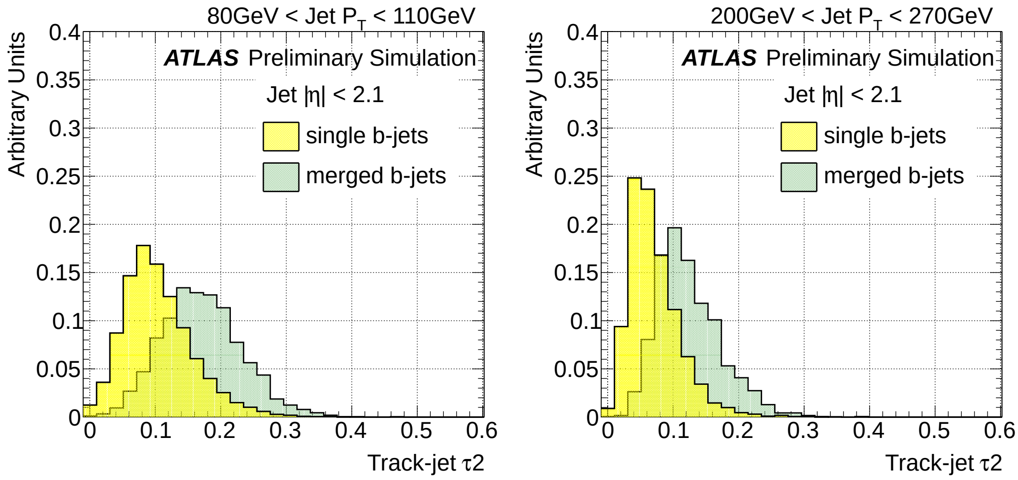


Figure 1.12: Distribution of τ_2 in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

more uniform track distribution. The distributions of τ_2, shown in Fig. 1.12, display good separation between single and merged jets, but with the latter showing larger values than single. This behavior can be traced to the level of correlation between τ_2 and track-jet width, displayed in Fig. 1.13a, to be compared to the much lower correlation presented, for instance, between track-jet width and the jet track multiplicity, shown in Fig. 1.13b.

The correlation observed suggests to switch from an absolute to a width-normalized τ_2, and evaluate the ratio τ_2/τ_1. Fig. 1.14 thus shows the distributions for this observable. Somewhat larger values are obtained for single than for merged b-jets, specially at high p_T, however we decided not to use this variable as it offers only marginal discrimination.

Jet eccentricity In defining a jet moment there are several ways to weight the momentum and define the center of the jet. We have defined the jet width as the first moment of the transverse energy with respect to the jet axis; another example of useful combination is the jet pull [17]. But it is

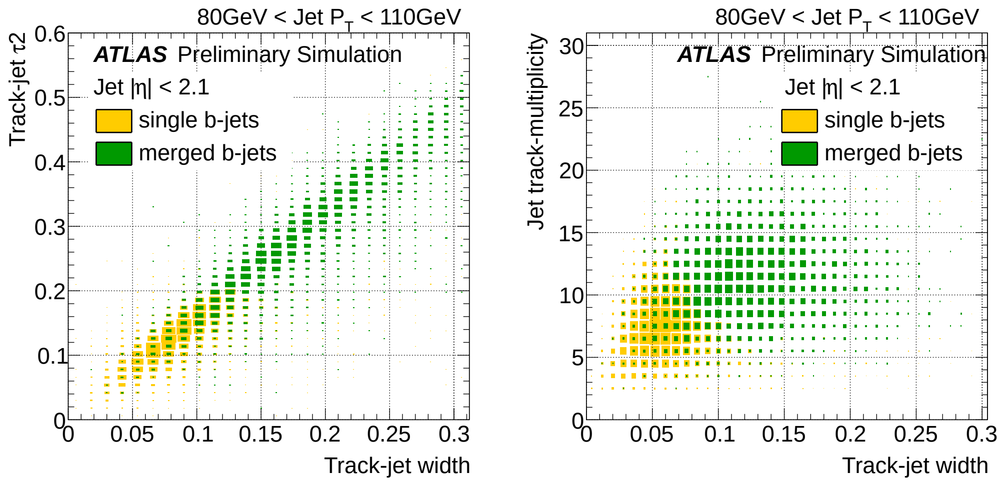


Figure 1.13: Correlation between τ_2 and track-jet width (left) and jet track multiplicity and track-jet width (right) for single and merged b-jets between 80 GeV to 110 GeV.

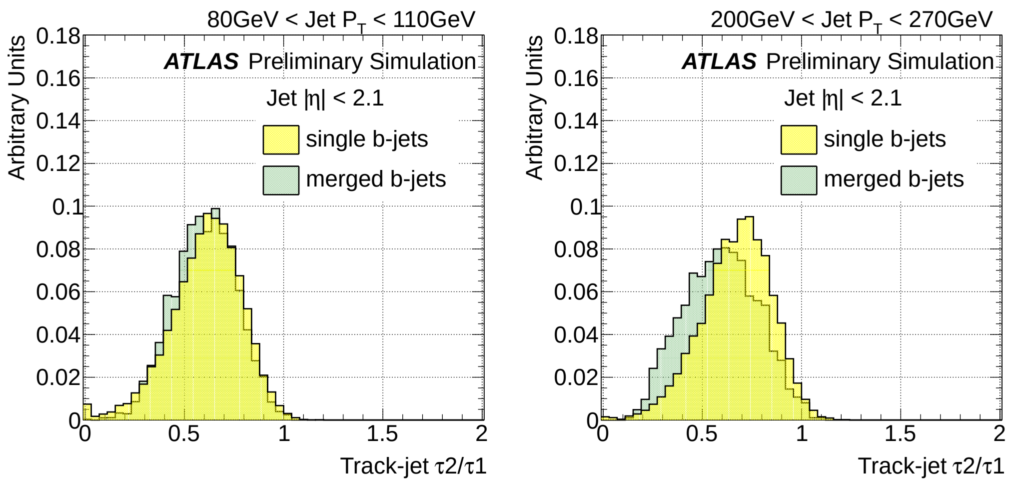


Figure 1.14: Distribution of τ_2/τ_1 in jets for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

also natural to look at higher moments, such as those contained in the 2×2 matrix,

$$\begin{bmatrix} \sum E_i \eta_i^2 & -\sum E_i \eta_i \phi_i \\ -\sum E_i \eta_i \phi_i & \sum E_i \phi_i^2 \end{bmatrix} \tag{1.5}$$

Here, (E_i, η_i, ϕ_i) are the jet constituent energy, pseudorapidity and azimuthal angle, respectively. The eigenvalues $\lambda_m \geq \lambda_p$ of this tensor are associated to the semiminor and semimajor axes of an elliptical jet, in the $\eta - \phi$ plane. The jet eccentricity, defined below, is a combination of these eigenvalues, and it is a measure of how elongated is the area of a jet,

$$e = \sqrt{1 - r^2} \tag{1.6}$$

where the parameter r is defined as the ratio of the eigenvalues,

$$r = \frac{\lambda_m}{\lambda_p} = \frac{\sum E_i \eta_i^2 + \sum E_i \phi_i^2 - \sqrt{(\sum E_i \eta^2 - \sum E_i \phi_i^2)^2 + 4(\sum E_i \eta_i \phi_i)^2}}{\sum E_i \eta_i^2 + \sum E_i \phi_i^2 + \sqrt{(\sum E_i \eta^2 - \sum E_i \phi_i^2)^2 + 4(\sum E_i \eta_i \phi_i)^2}}. \tag{1.7}$$

Figure 1.15 shows the distribution of the jet eccentricity, built using track constituents. No significant difference in eccentricity was found between single and merged b-jets.

Further studies

In order to better understand the behavior observed for τ_2, τ_2/tau_1 and jet eccentricity in anti-k_T 0.4 jets, these variables were studied for other two different scenarios,

- using the active area of jets (with clusters used as input to jet reconstruction).
- using bigger 0.6 anti-k_T jets

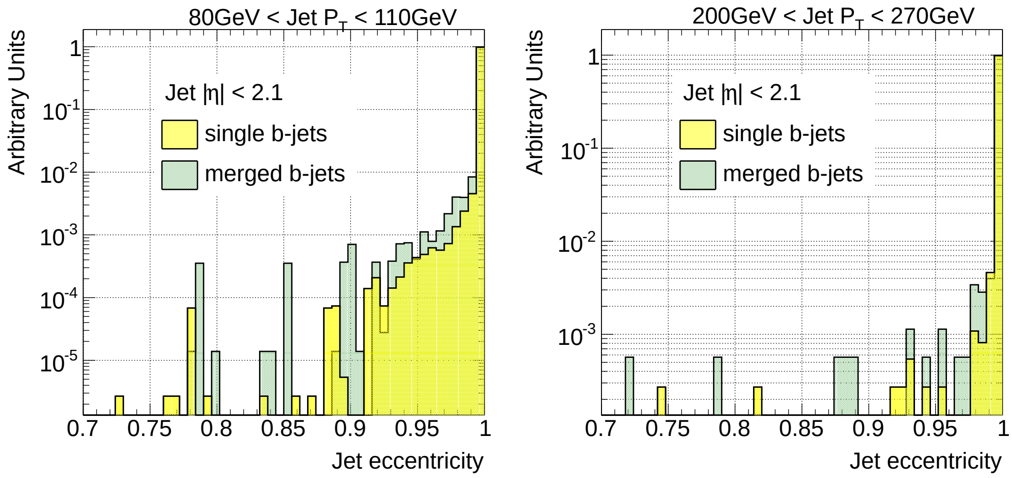


Figure 1.15: Distribution of the jet eccentricity for single and merged b-jets between 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

in order to enhance the efficiency to capture the decay products in gluon to $b\bar{b}$-jets.

Figures 1.17 to 1.18 show distributions of variables mentioned above for single and merged b-jets between 80 GeV to 110 GeV.

1.3 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 of comparison plots is presented. Figures 1.19 and 1.20 show distributions of jet track multiplicity, track-jet width, ΔR between the axes of the two k_t subjets, $\max\{\Delta R(trk, trk)\}$ and τ_2 in two different p_T bins for b-tagged jets in di-jet Monte Carlo and data events passing selection described in Section 1.1.1. The distributions are normalized to unit area to allow for shape comparisons. There is a good agreement between data and simulation.

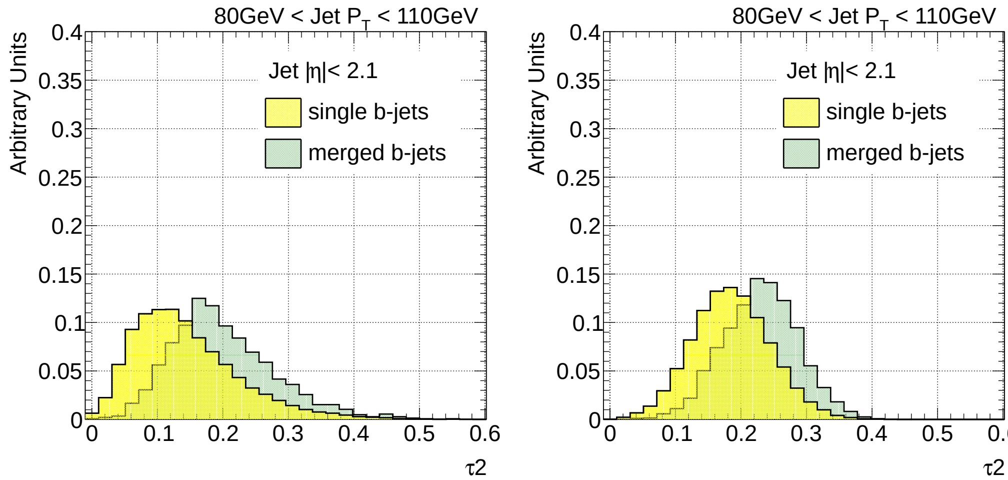
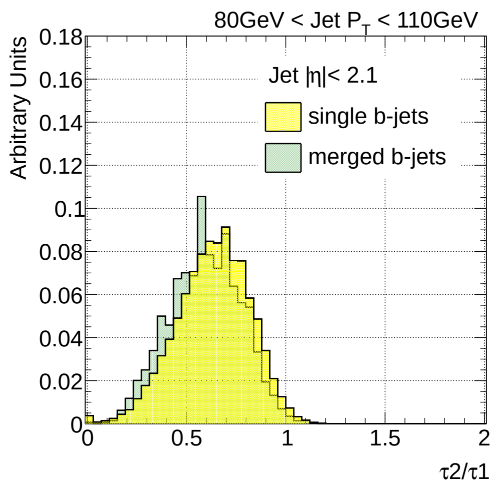


Figure 1.16: Distribution of τ_2 for single and merged b-jets between 80 GeV to 110 GeV in anti-k_T 0.6 jets using track constituents (left) and anti-k_T 0.4 jets using the active area of the jet, with calorimeter topoclusters as input.



FIGS/TEMPFigs/GhostMatchingVarsClus/Tauratio2080

Figure 1.17: Distribution of τ_2/τ_1 for single and merged b-jets between 80 GeV to 110 GeV in anti-k_T 0.6 jets using track constituents (left) and anti-k_T 0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

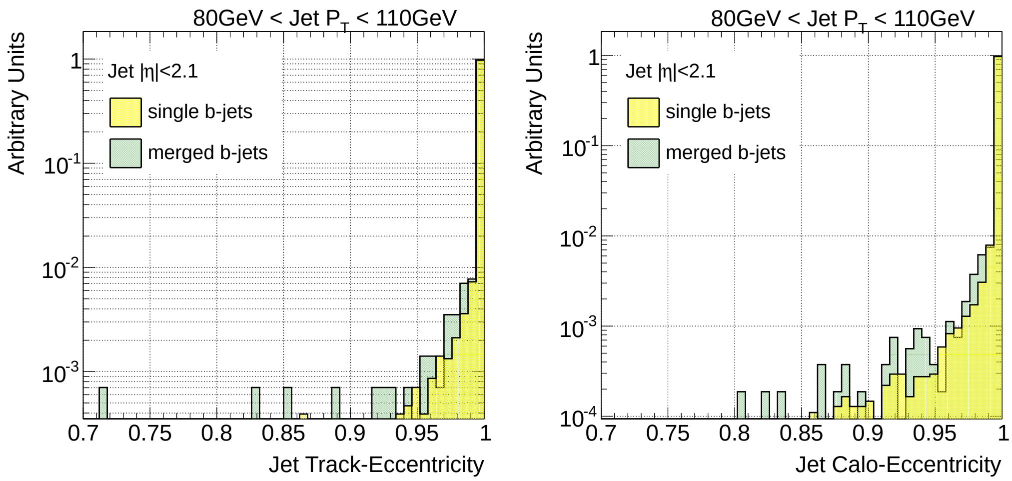


Figure 1.18: Distribution of the jet eccentricity for single and merged b-jets between 80 GeV to 110 GeV in anti-k_T 0.6 jets using track constituents (left) and anti-k_T 0.4 jets using the active area of the jet, with calorimeter topoclusters as input.

It should be remarked that the observed agreement is actually not a direct validation of the description in the MC of the relevant variables, but its convolution with the simulated relative fractions of light-, c-, b- and bb-jets in the b-tagged generated jet sample. To some extent, there could be some level of compensation between these two effects, although the agreement evaluated in b-jets selected with a looser cut of MV1 tagger as well as with another b-tagging algorithm is still ver good, suggesting that this compasation is not likely to occur in samples sufficiently enriched in b-jets.

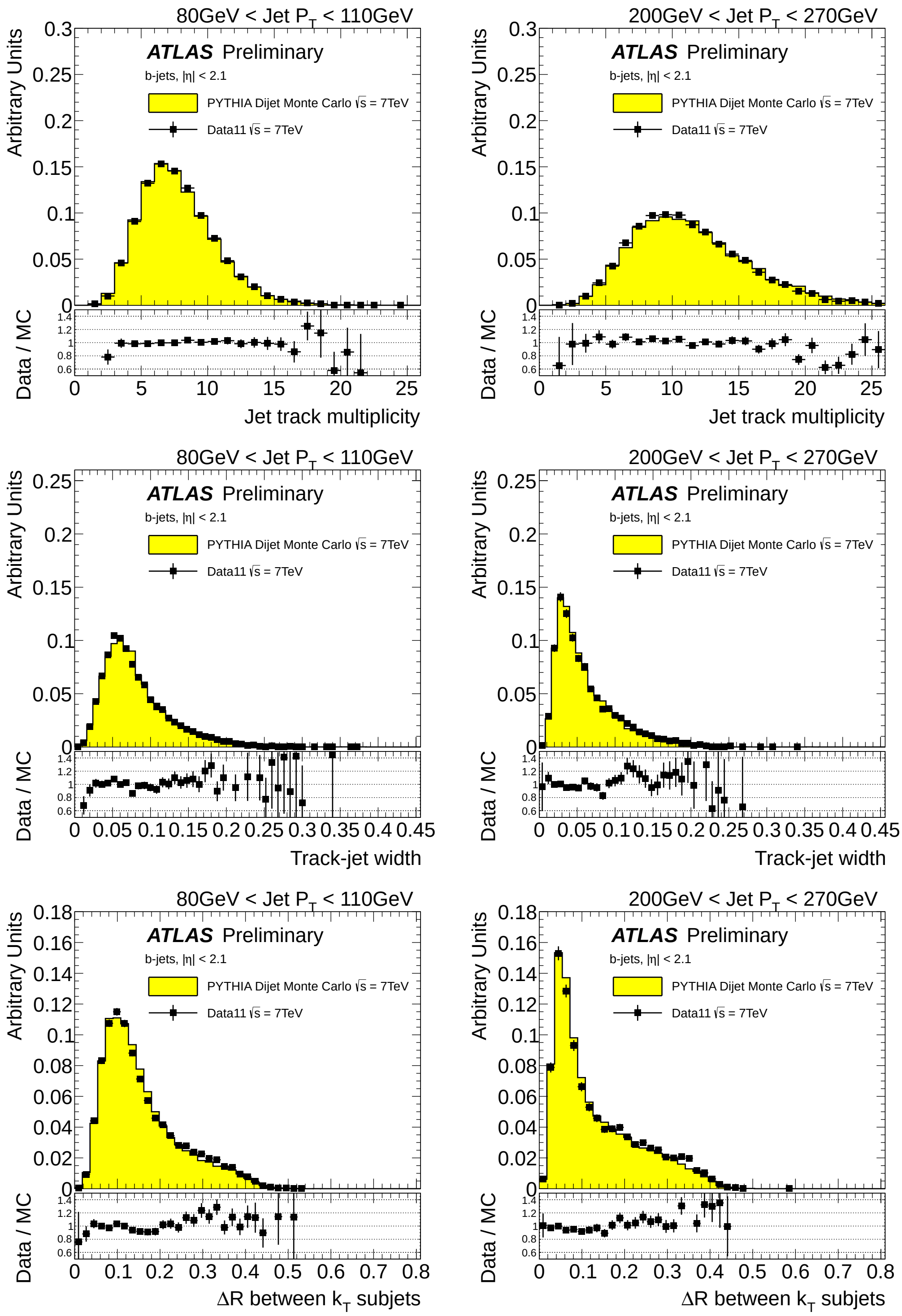


Figure 1.19: Distribution of three tracking variables in 2 different jet p_T bins, for experimental data collected by ATLAS during 2011 (solid black points), and simulated data (filled histograms). The ratio data over simulation is shown at the bottom of each plot.

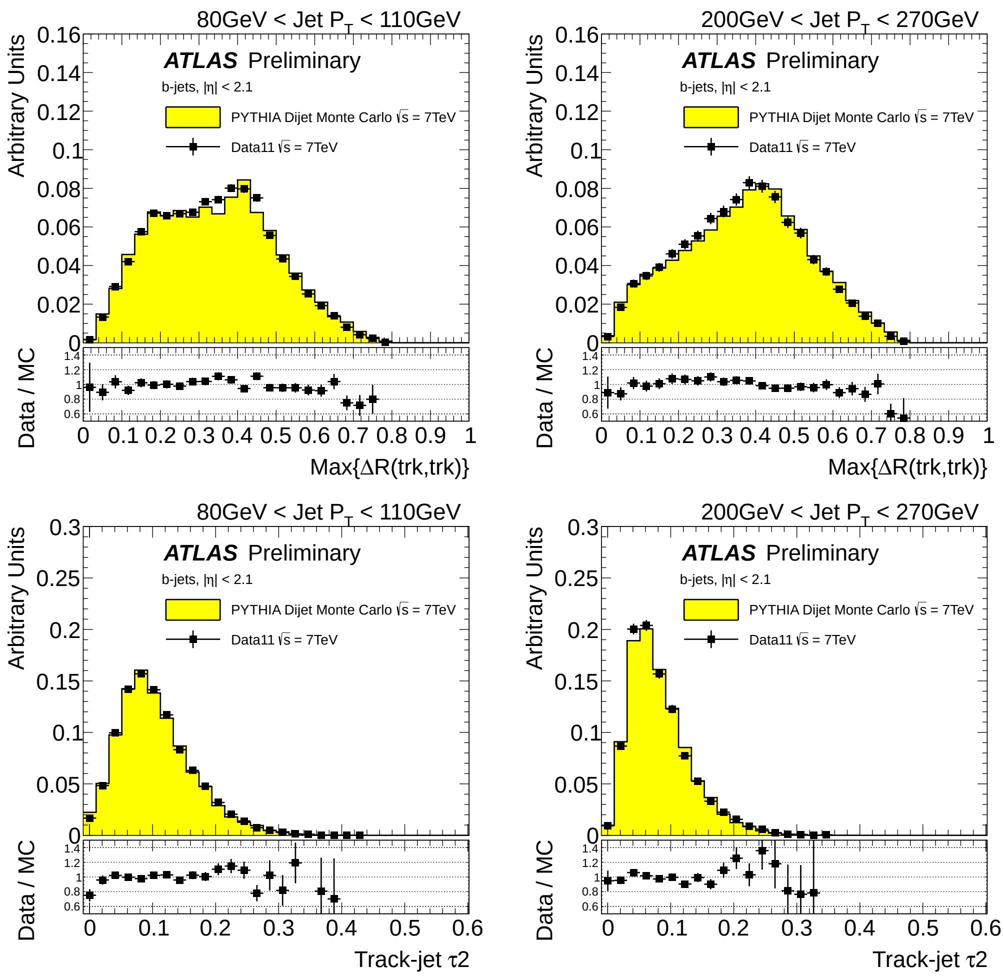


Figure 1.20: Distribution of two tracking variables in two different jet p_T bins, for experimental data collected by ATLAS during 2011 (solid black points), and simulated data (filled histograms). The ratio data over simulation is shown at the bottom of each plot.

Bibliography

[1] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[2] Atlas tunes of pythia 6 and pythia 8 for mc11. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011.

[3] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

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

[5] ATLAS Collaboration. Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data. *ATLAS-CONF-2012-020*, 2012.

[6] Jason Gallicchio and Matthew D. Schwartz. Quark and gluon tagging at the lhc. *Phys. Rev. Lett.*, 107:172001, Oct 2011.

[7] ATLAS Collaboration. Light-quark and gluon jets in atlas. *ATLAS-CONF-2011-053*, 2011.

[8] D. Buskulic et al. Quark and gluon jet properties in symmetric three-jet events. *Physics Letters B*, 384(1âĂŞ4):353 – 364, 1996.

[9] Leandro G. Almeida, Seung J. Lee, Gilad Perez, Ilmo Sung, and Joseph Virzi. Top quark jets at the lhc. *Phys. Rev. D*, 79:074012, Apr 2009.

[10] R. Snihur. Subjet multiplicity in quark and gluon jets at d0. *Nuclear Physics B - Proceedings Supplements*, 79(1âĂŞ3):494 – 496, 1999. <ce:title>Proceedings of the 7th International Workshop on Deep Inelastic Scattering and QCD</ce:title>.

[11] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.

[12] A. Abdesselam, E. Bergeaas Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, et al. Boosted objects: A Probe of beyond the Standard Model physics. *Eur.Phys.J.*, C71:1661, 2011.

[13] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Techniques for improved heavy particle searches with jet substructure. *Phys. Rev. D*, 80:051501, Sep 2009.

[14] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, Oct 1993.

[15] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 1103:015:026, 2011.

[16] Iain W. Stewart, Frank J. Tackmann, and Wouter J. Waalewijn. n jettiness: An inclusive event shape to veto jets. *Phys. Rev. Lett.*, 105:092002, Aug 2010.

[17] Jason Gallicchio and Matthew D. Schwartz. Seeing in color: Jet superstructure. *Phys. Rev. Lett.*, 105:022001, Jul 2010.