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

María Laura González Silva

Doctoral Thesis in Physics

Physics Department

University of Buenos Aires

November 2012

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

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

Trabajo de Tesis para optar por el título de
Doctor de la Universidad de Buenos Aires en el área Ciencias Físicas

por **María Laura González Silva**

Director de Tesis: Dr. Ricardo Piegaia
Lugar de Trabajo: Departamento de Física

Buenos Aires, Noviembre 2012

Agradecimientos

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

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

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

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

Resumen

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

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

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

Abstract

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

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

Contents

Chapter 1

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 sample

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 (see Fig.**??**). This fact will be of importance when discussing the selection of discriminating variables.

The events were 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-p_T 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.

1.2 Monte Carlo sample

The Monte Carlo samples employed were produced with the event generators discussed in Section **??**. 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 $2 \rightarrow 2$ 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, seven samples were generated with different thresholds of the hard-scattering partonic transverse momentum $\hat{p}_T$: 8-17 GeV, 17-35 GeV, 35-70 GeV, 70-140 GeV, 140-280 GeV, 280-560 GeV and 560-1120 GeV. For the Monte Carlo p_T distribution, or the distribution of any other observable, to be compared to that in experimental data, events from the different samples need to be weighted by their respective production cross sections. The p_T pectrum obtained after performing this procedure is displayed in Fig. 1.1.

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.

1.2.1 Event and jet selection

The data sample in the analysis is selected online using a set of single jet triggers as described in Section 1.1. In the case of the Monte Carlo, a trigger simulator is used. In this way both the simulated and real data from the detector can then be run through the same ATLAS trigger packages [3].

The offline event selection comprises an additional set of cuts on the reconstructed objects, including jet kinematic and jet-specific data quality

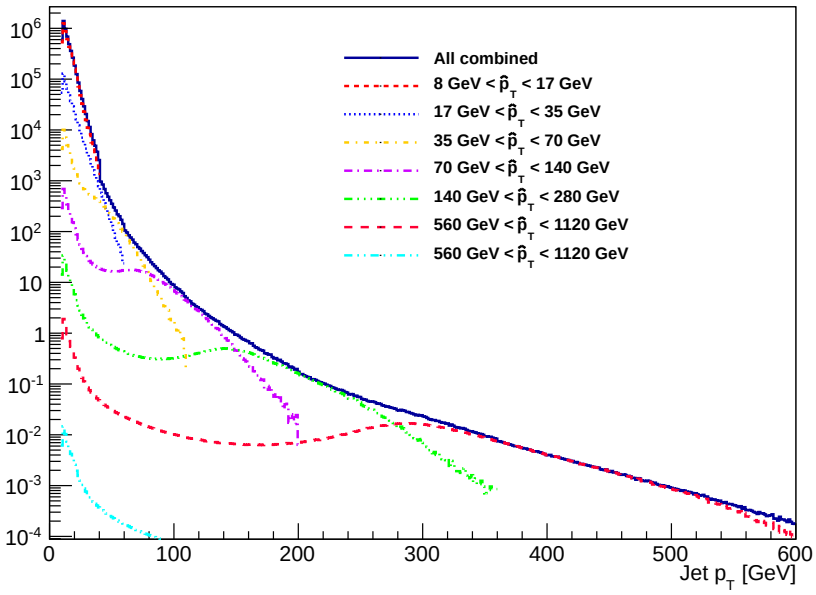

Figure 1.1: Calibrated jet p_T distribution for anti-k_t jets in a dijet Monte Carlo sample composed of seven sub-samples generated with different thresholds of the hard-scattering partonic transverse momentum, $\hat{p}_T$. In order to obtain the falling p_T spectrum observed in data, the different samples were weighted by their respective production cross sections.

cuts. A vertex cut is also included, requiring at least one primary vertex with five or more associated tracks in the event. This cut serves as a first rejection for events originating from cosmic rays and particles produced in interactions of the beam with particles in the beam tunnel ("beam halo" and "beam gas"). 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.

The jet algorithm selected for the analysis was the ATLAS default anti-k_t algorithm (Section **??**), with a distance parameter $R = 0.4$, using calorimeter topological clusters as input (Section **??**). All jets were calibrated using the EM+JES scheme (Section **??**). A high cut on the minimum jet p_T is implemented to select jets in the region where the triggers used in the analysis are most efficient. Jets are required to have a minimum p_T of 40 GeV. Jets with transverse momentum above this threshold were also required to be in a region with full tracking coverage, $|\eta_{jet}| < 2.1$. Although the Pixel and SCT detectors cover up to $|\eta| < 2.5$, a lower pseudorapidity cut is used in order to account for the size of the calorimeter jets, $R = 0.4$. Jets passing this selection 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. The upper limit of our highest p_T bin was set to 480 GeV; beyond this energy the b-tagging efficiency becomes very poor.

Several quality criteria are applied to jets to eliminate "fakes" that are caused by noise bursts in the calorimeters and energy depositions belonging to a previous bunch crossing. A detailed description of these quality cuts can be found in reference [4].

In addition to these kinematic and quality cuts, two more cuts are imposed to jets:

- ***b*-tagging**. Jets are only accepted if they are tagged as *b*-jets using the MV1 *b*-tagging algorithm, at its 60% efficiency working point.

- **Isolation**. Jets are only accepted if they are isolated. The isolation criterion requires that no other jet with a $p_T > 7$ GeV be within $\Delta R < 2R$, where R is the distance parameter of the jet algorithm.

Finally, 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 contained two (only one) *b*-hadron:

$$\text{single } b\text{-jets: } \Delta R(j, B_i) < 0.4 \;\&\; \Delta R(j, B_j) > 0.4 \;\; for\; i \neq j \tag{1.1}$$

$$\text{merged } b\text{-jets: } \Delta R(j, B_i) < 0.4 \;\&\; \Delta R(j, B_j) < 0.4 \;\; for\; i \neq j \tag{1.2}$$

where j is a jet in the event and $B_{i(j)}$ 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.

1.2.2 Track selection

The tracking system provides a very precise tool for understanding the structure of jets and for mitigating the pile-up background. Charged particle jet constituents that leave tracks in the inner detector provide 3-dimensional information on the jet origin and direction as a result of the vertexing provided by the tracks. The combination of tracking and calorimetry therefore

greatly enhance the identification and selection of hadronic jets from primary interactions that do typically have associated charged tracks.

In the study of the internal structure of jets containing b-hadrons, the tracking information will be used to define jet variables with potential discriminating power between single and merged b-jets. For this reason the selection of genuine tracks belonging to jets is one of the most important steps of the analysis.

The jet direction is used to associate the charged particles reconstructed as tracks in the inner detector to the jet. A simple $\Delta R < 0.4$ matching criterion is used, where the matching is performed using the track coordinates at the point of closest approach to the primary vertex.

Tracks are required to fulfill cuts on their transverse momentum, number of hits and transverse and longitudinal impact parameters, similar to those applied by b-tagging algorithms (see Section **??**). Cuts on ${p_T}^{trk} > 1.0$ GeV and the χ^2 of the track fit, $\chi^2/ndf < 3$, are applied. The effect of a lower cut on the track transverse momentum, ${p_T}^{trk} > 1.0$ GeV, is discussed in the next section. 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. As cutting on impact parameter (IP) might be detrimental for b-jets, where large IP values are expected, relaxed cuts were used, $|d_0| <$ 2 mm, and $|z_0 \sin\theta| < 2$ 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 tracks associated to b-jets, 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.

1.3 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 1.2.2, and contained within a cone of radius $R = 0.4$ around the jet axis. Figure 1.2 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.

We also considered the possibility of restricting ourselves to using tracks significantly displaced from the PV ($|d_0|/\sigma(d_0) > 2.5$), which are more likely to originate from the b-hadrons decays. In order to evaluate the effect of this particular selection, a preliminary study was done with a sample of dijet events generated with PYTHIA and with no detector simulation (denoted as "standalone" PYTHIA in the following). 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, the equivalent of tracks at the level of event generation, was done in the same way as with the full ATLAS simulation. Distributions of the track multiplicity built using all charged

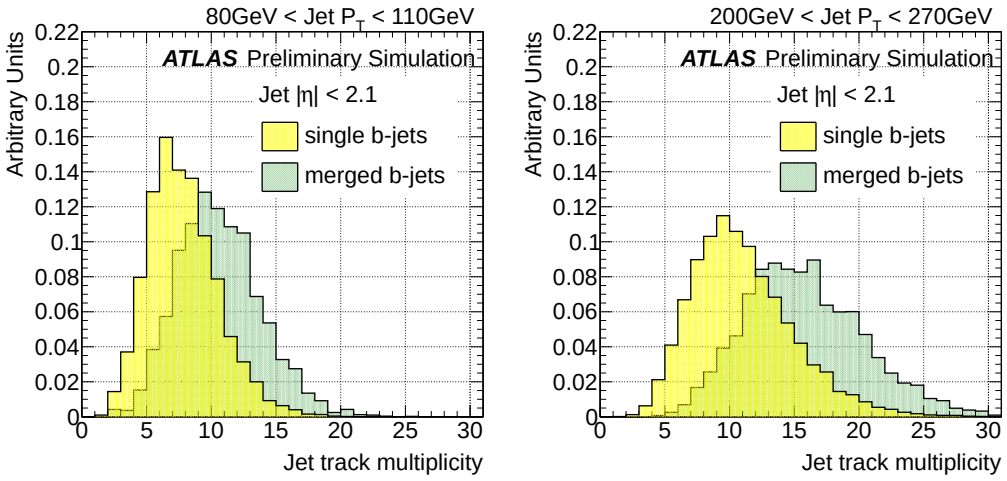

Figure 1.2: Distribution of the track multiplicity in jets for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

particles and using only charged particles coming from the b-hadron decay ("b-tracks") are illustrated in Fig. 1.3. A better discrimination between single and merged b-jets, measured in terms of the significance, $s = \Delta \mathrm{n}_{trk}/\sigma(\Delta \mathrm{n}_{trk})$ with n_{trk} the mean jet track multiplicity, is observed when using b-tracks only: $s = 5.9 \cdot 10^{-1}$ compared to $s = 4.4 \cdot 10^{-1}$ when using all charged particles. The result obtained with standalone PYTHIA suggests that a potential improvement in single-merged separation can be achieved by circumscribing the track selection, in the full simulation, to tracks with large impact parameter significance. A comparison of track multiplicity distributions using all tracks and distributions built with displaced tracks only is shown in Fig. 1.4. No improvement is obtained by using displaced tracks. The potential sensitivity achieved by enriching the sample in tracks associated to the b-hadron is counterbalanced by the lower number of associated tracks.

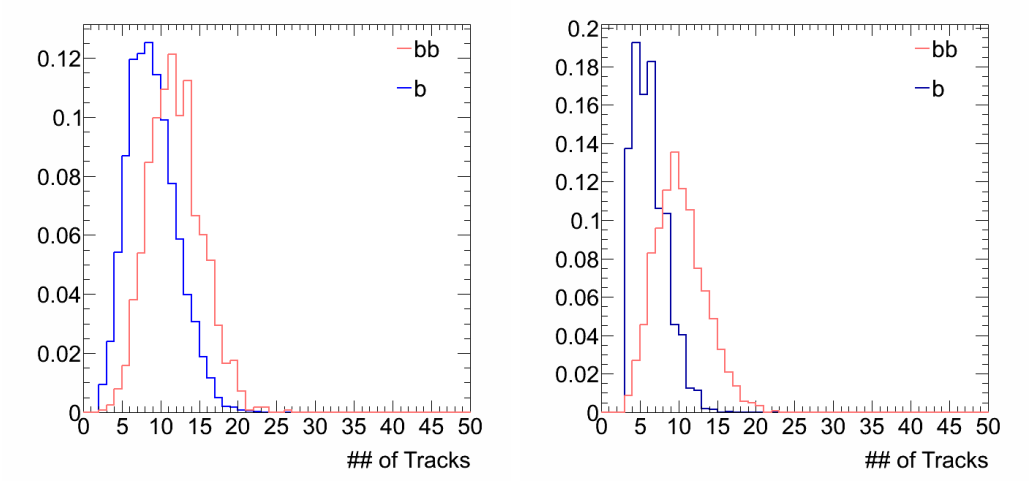

Figure 1.3: Distribution of the charged particle multiplicity for single (b) and merged (bb) jets from 80 GeV to 120 GeV in a sample of dijet events generated with PYTHIA and no detector simulation. Distributions are shown using all charged particles (left) and using only charged particles coming from b-hadron decay (right). A better discrimination between single and merged b-jets is obtained when using tracks from b-decay only.

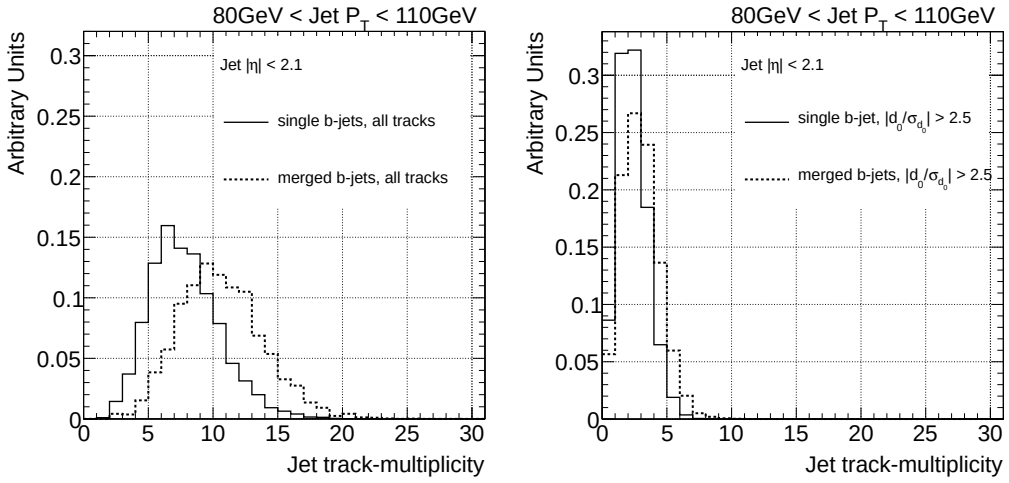

Figure 1.4: Distribution of the jet track multiplicity for single and merged b-jets from 80 GeV to 110 GeV, for all (left) and displaced tracks only (right). No improvement is obtained by using displaced tracks.

Jet width

The jet width is part of a set of continuous variables, like geometric moments, that are sensitive to the distribution of the constituents within a jet. This particular combination is a linear moment which sums the distances between the jet constituents and its axis, weighted by the constituents p_T. 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, track or particle constituents.

This observable has also found use in the discrimination between gluon initiated and light quark initiated jets, see for instance [5] and [6]. Gluon jets 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 [7]: due to the mass diference between b-hadrons and light-quark 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 1.5 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 b-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, with more jet-to-jet fluctuations, while the particle multiplicity is relatively fixed in the b-hadron decay.

The distribution of the track-jet width for the full ATLAS simulation is shown in Fig. 1.6. 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 used to 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.

In full simulation, the jet width can be measured in terms of calorimeter variables by replacing tracks by topological clusters in the sum (this is somehow the equivalent in full simulation of switching from charged to all particles). 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. 1.7, 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

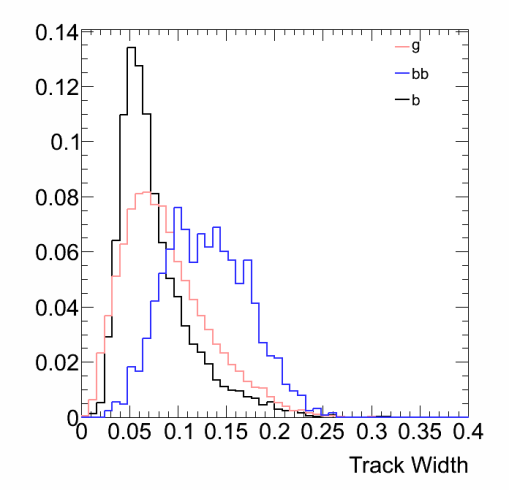

Figure 1.5: Distribution of track-jet width in jets for gluon-initiated (g), single (b) and merged (bb) jets from 80 GeV to 120 GeV in a sample of dijet events generated with PYTHIA and no detector simulation.

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, depends on the radiation pattern of the event. It is the most basic observable for distinguishing massive boosted objects from jets originating from quarks or gluons. The latter are expected to be dominated by wide-angle emissions, with increased probability to see high mass jets initiated from gluons as opposed to quarks [8].

Detector level jet mass distributions for jets selected to have $80 < p_T < 110$ GeV and $200 < p_T < 270$ GeV are shown in Fig. 1.8, both for single and merged b-jets. Merged jets tend to have higher masses than single b-jets for

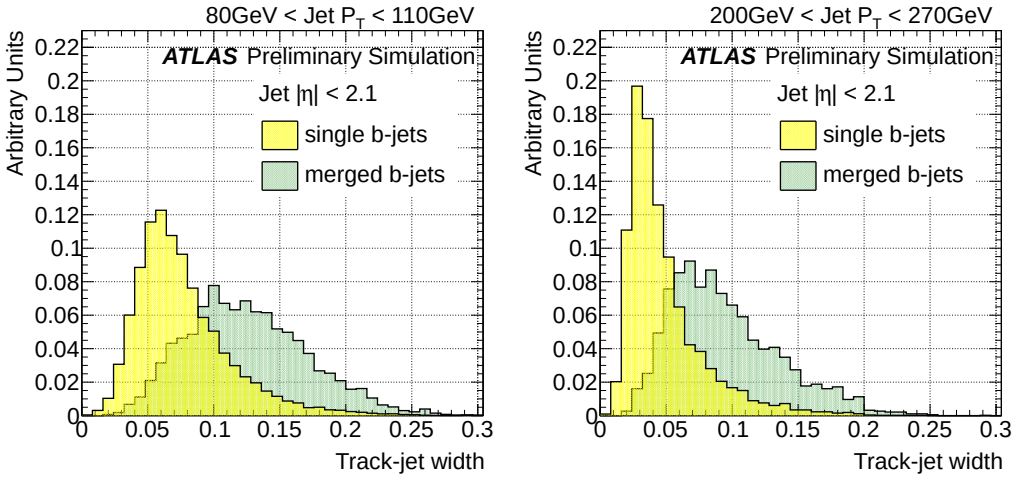

Figure 1.6: Distribution of track-jet width for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

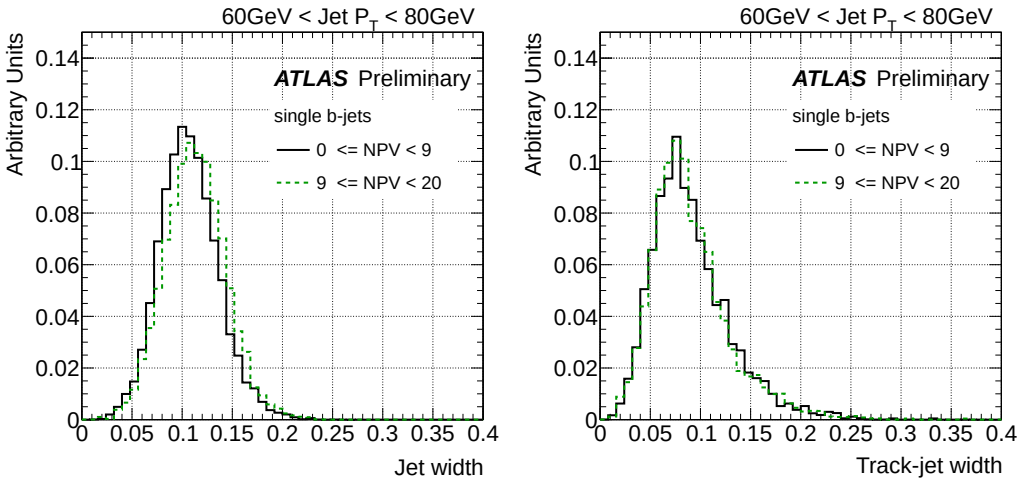

Figure 1.7: Distribution of jet width using topological clusters (left) and tracks (right) for single b-jets in two bins of number of primary vertices (NPV) for jets from 60 GeV to 80 GeV.

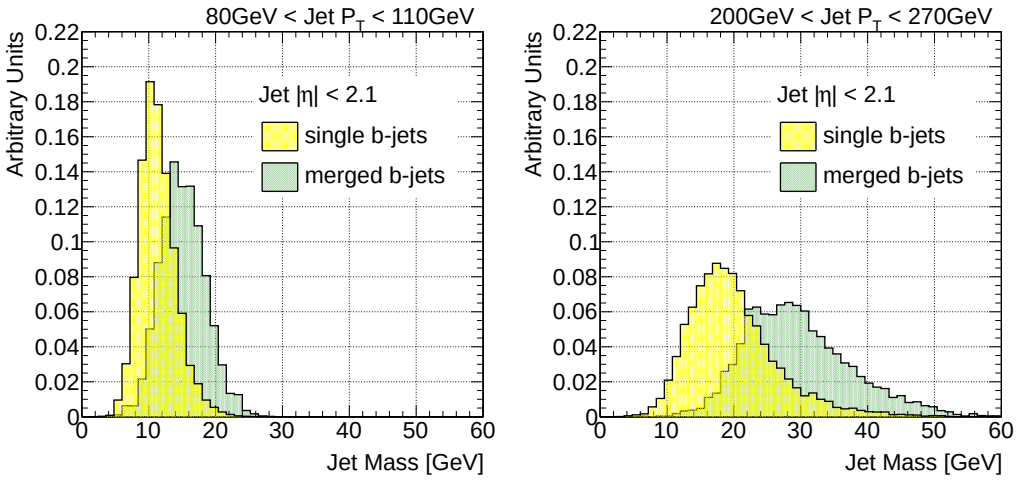

Figure 1.8: Distribution of jet mass in GeV for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

the same p_T bin. Although it shows good separation, this calorimeter based variable can be significantly affected by the amount of pile-up in the event as even a single soft wide angle deposition will have an effect on the jet mass, shifting the distribution to higher values[1].

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 1.9 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

[1]In the ATLAS analysis of 35 pb^{-1} of 2010 data, the sensitivity of individual jet mass to pile-up is directly tested (for jets with at least 300 GeV). The mean jet mass is observed to increase linearly with NPV [9].

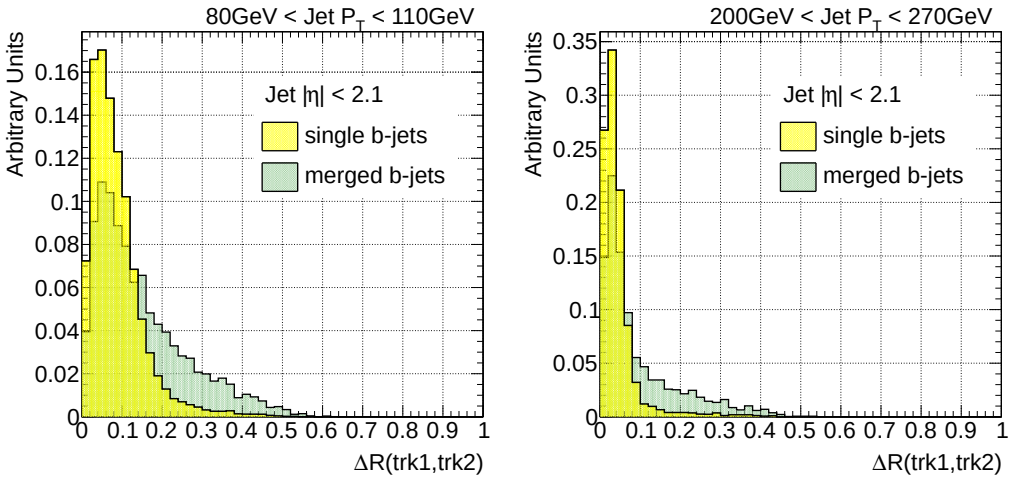

Figure 1.9: Distribution of ΔR between leading tracks for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

broader than single b-jet distributions for medium jet 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 ($\max\{\Delta R(trk, trk)\}$) is one example. Its distribution is shown in Fig. 1.10, for single and double b-hadron jets. The latter show significantly higher values 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, alternative characterising variables are desirable as $\max\{\Delta R(trk, trk)\}$ is sensitive to soft tracks originating from pile-up.

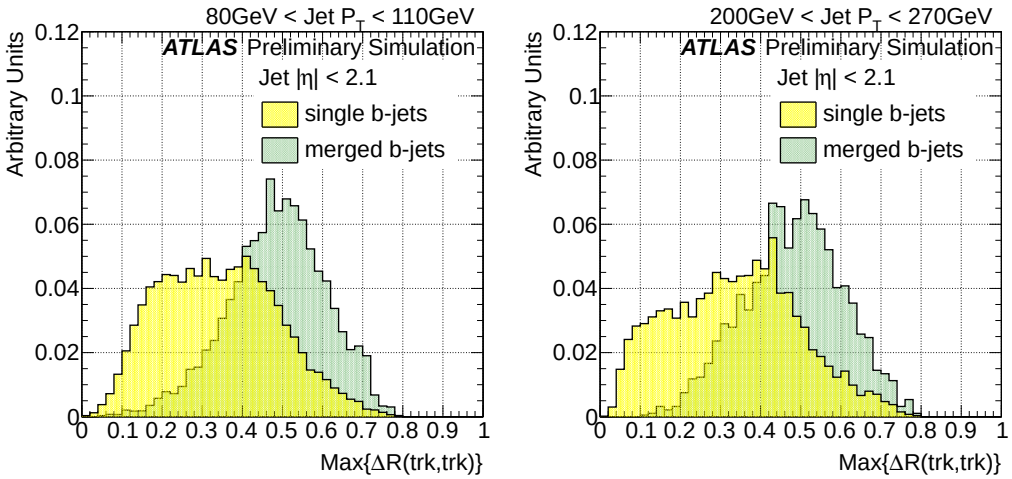

Figure 1.10: Distribution of the maximum ΔR between pairs of tracks in jets for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

Subjet multiplicity

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, a measurement of this observable for quark and gluon jets indicates that gluon-initiated jets tend to have on average higher subjet multiplicity [10]. This result is consistent with the QCD prediction that gluons radiate more than quarks. In this case the k_t algorithm is rerun for subjet finding.

In general, by using the inclusive k_t or other sequential recombination algorithms, introduced in Section **??**, it is straightforward to define a "subjet algorithm" in which the structure of the jet 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

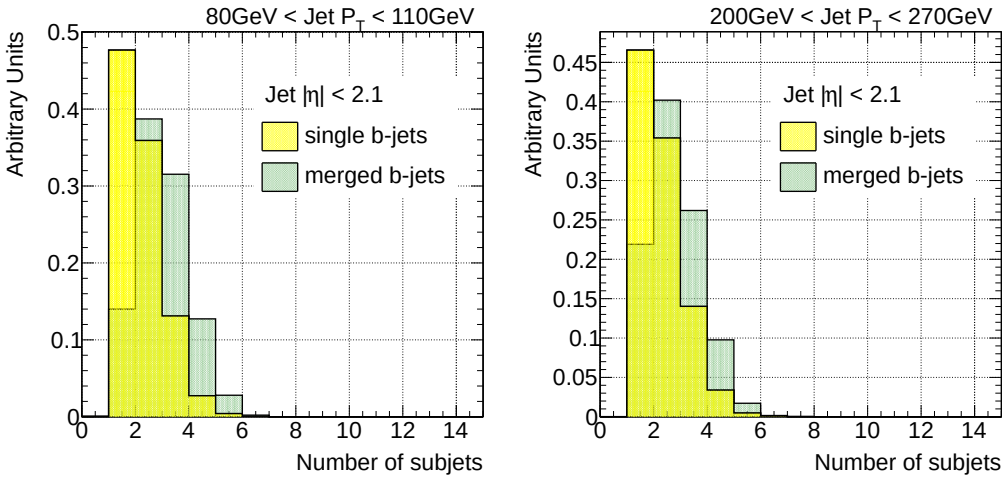

Figure 1.11: Distribution of the number of k_t sub-track-jets for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

recombination sequence [11] in order to identify the decay products of an object. This approach is used in several jet grooming procedures[2], see for instance [13].

Figure 1.11 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 $R = 0.2$. Merged jets tend to have on average one more subjet than single b-jets. The discrimination power of this variable is very poor.

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

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

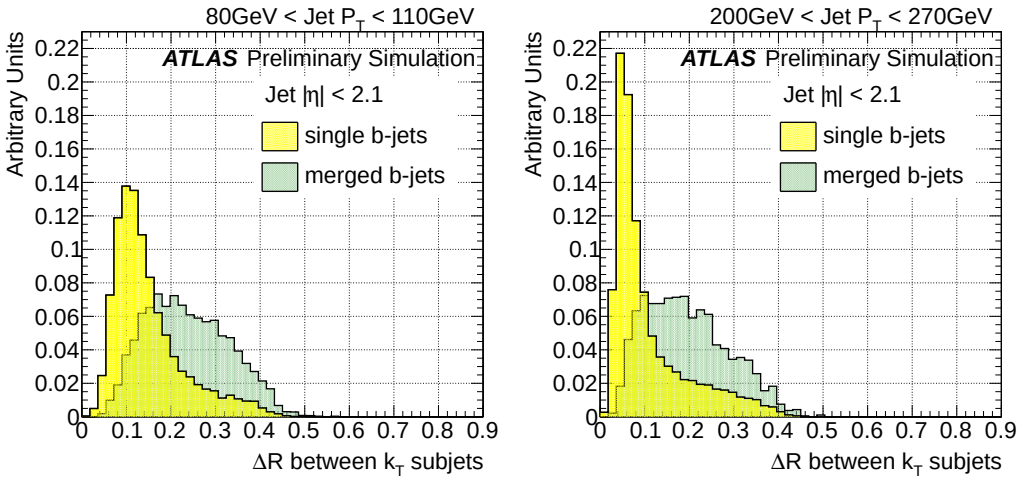

Figure 1.12: Distribution of the ΔR between the axes of the two k_t subjets in the jet for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

once it reaches exactly two jets. The ΔR between the axes of the two exclusive subjets is shown in Fig. 1.12. 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 $\max\{\Delta R(trk, trk)\}$.

In order to illustrate what this variable represents, an event display of a merged b-jet with a large (> 0.3) ΔR value is shown in Fig. 1.13. For this case, the jet active area [15] was reconstructed following a procedure called "ghost association", in which the event topoclusters plus a grid of 0.1×0.1 cells with nearly zero energy get associated to the jet through jet clustering. In the case of Figure 1.13 the anti-k_t algorithm was used. The event display indicates how the high energy cells in the jet with two b-hadorns are grouped around the b-hadrons directions, leading to the two-subjet substructure of merged jets.

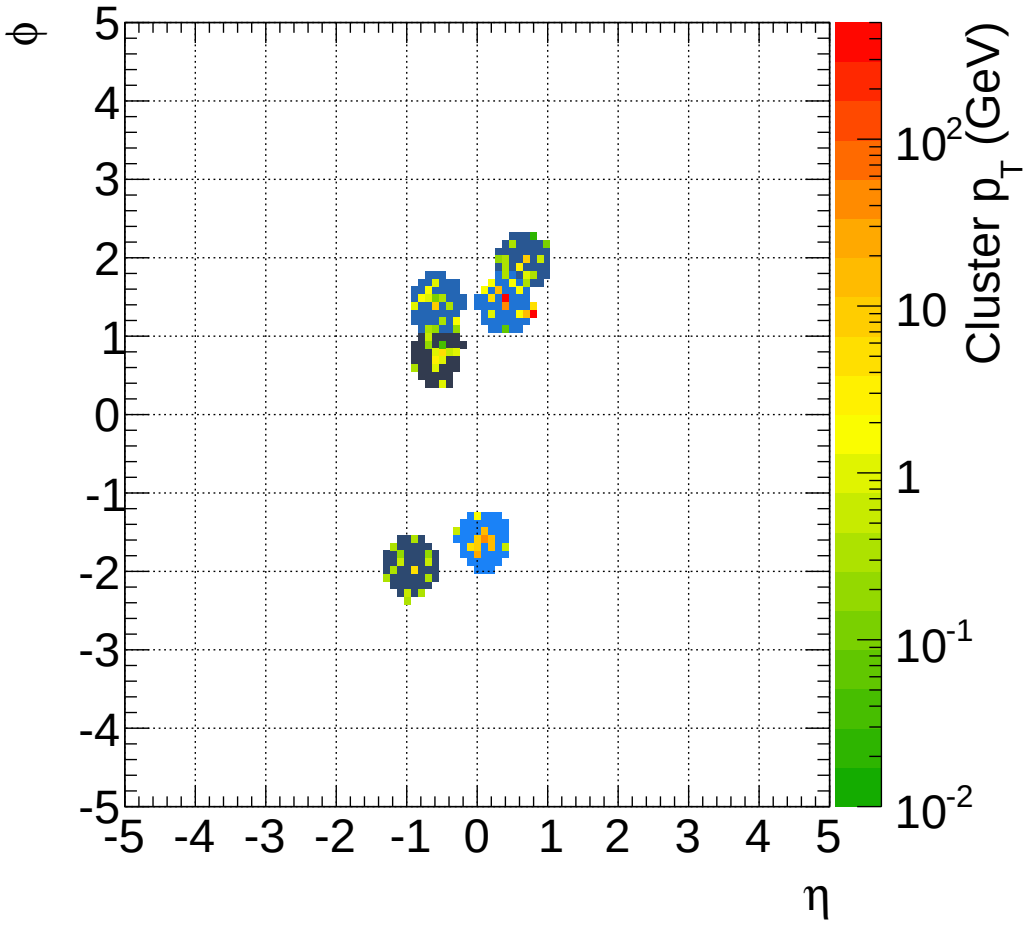

Figure 1.13: Event display of a merged b-jet in $(\eta, \phi) = (0.46, 1.41)$ and $p_T = 110$ GeV. The two b-hadrons are indicated as two red squares. The active area of the jet is shown in blue and the topoclusters belonging to the jet are shown in different colors, from green to orange, depending on their transverse momentum. The ΔR between the axes of the two k_t subjets in the jet is larger than 0.3. The two-subjet structure of the merged jet is displayed.

N-subjettiness variables

It is possible to extend the use of individual subjets in conjunction with more sophisticated jet shape variables. Using these tools, an inclusive jet shape based on the substructure topology of a single jet, "N-subjettiness" has been recently proposed [16]. This variable describes the energy flow within a jet, quantifying the degree to which radiation is aligned along N subjet axes. That is, it characterizes how consistent a jet is with an N-subjet substructure. This jet shape was adapted from the event shape N-jettiness [17].

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 constituents 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 the p_T of the constituent particle, this variable is an infrared-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 [16] . 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 choices 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 more uniform track distribution. The distributions of τ_2, shown in Fig. 1.14, 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. 1.15a, to be compared to the much lower correlation presented, for instance, between track-jet width and jet track multiplicity, shown in Fig. 1.15b.

The correlation observed suggests to switch from an absolute to a width-normalized τ_2, and evaluate the ratio τ_2/τ_1, as shown in Fig. 1.16. 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 [18]. But it is also natural to look at

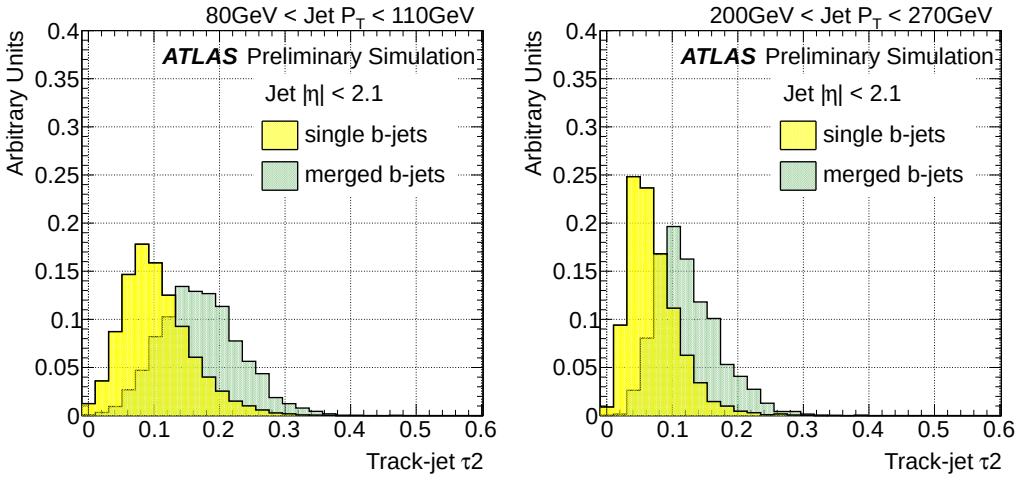

Figure 1.14: Distribution of τ_2 for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

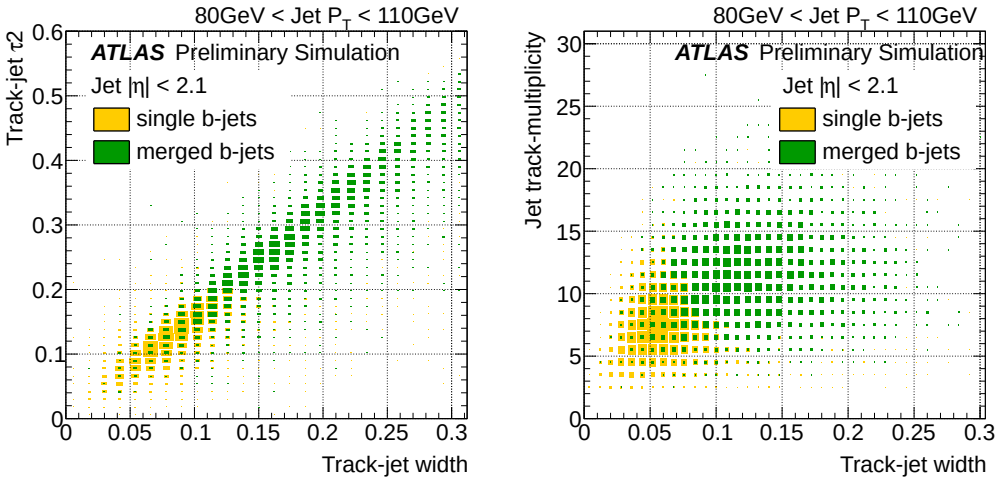

Figure 1.15: Correlation between τ_2 and track-jet width (left) and jet track multiplicity and track-jet width (right) for single and merged b-jets from 80 GeV to 110 GeV.

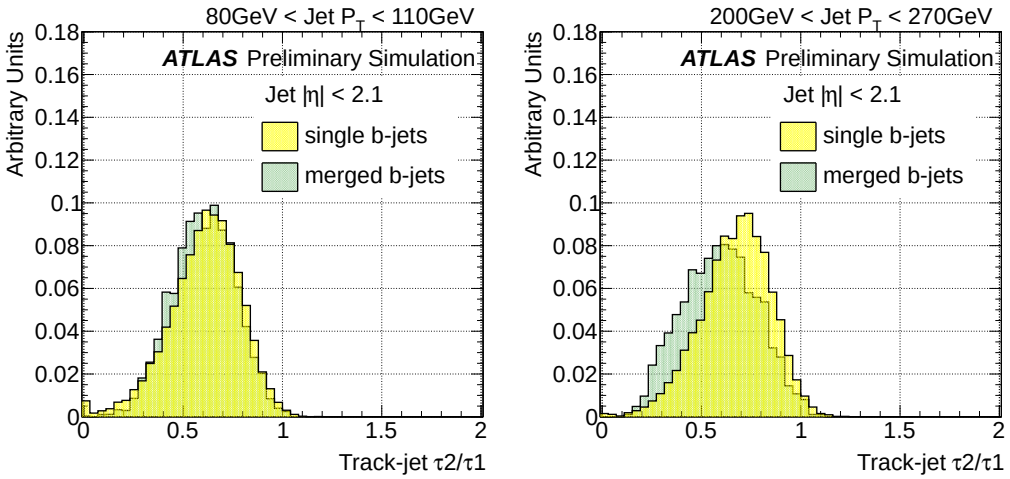

Figure 1.16: Distribution of τ_2/τ_1 for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

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 $\lambda_m \leq \lambda_p$ of this tensor are associated to the semiminor and semimajor axes of an elliptical approximation to the jet shape in the $\eta - \phi$ plane. The jet eccentricity, defined below, is a combination of these eigenvalues, and it is a measure of how elongated the area of a jet is,

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

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 1.17 shows the distribution of the jet eccentricity, built using track constituents. No significant difference in eccentricity was found between sin-

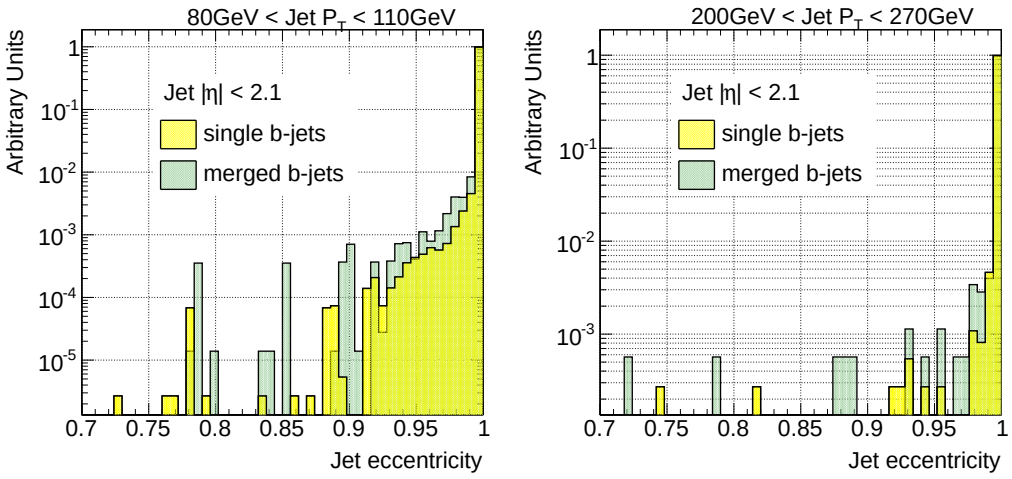

Figure 1.17: Distribution of the jet eccentricity for single and merged b-jets from 80 GeV to 110 GeV (left) and 200 GeV to 270 GeV (right).

gle and merged b-jets.

In order to better understand the behavior observed for this variable, further studies were performed using bigger cone jets to enhance the efficiency to capture the decay products in gluon to $b\bar{b}$-jets, in particular the anti-k_t 0.6 algorithm was used; and the ghost association procedure for cluster matching, for a calorimeter-based eccentricity. None of the above resulted in better single-merged separation.

1.4 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 1.18 and 1.19 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 dijet

Monte Carlo and data events passing the selection described in Section 1.2.1. The distributions are normalized to unit area to allow for shape comparisons. There is a very good agreement between data and simulation in all cases.

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 very good, suggesting that this compensation is not likely to occur in samples sufficiently enriched in b-jets.

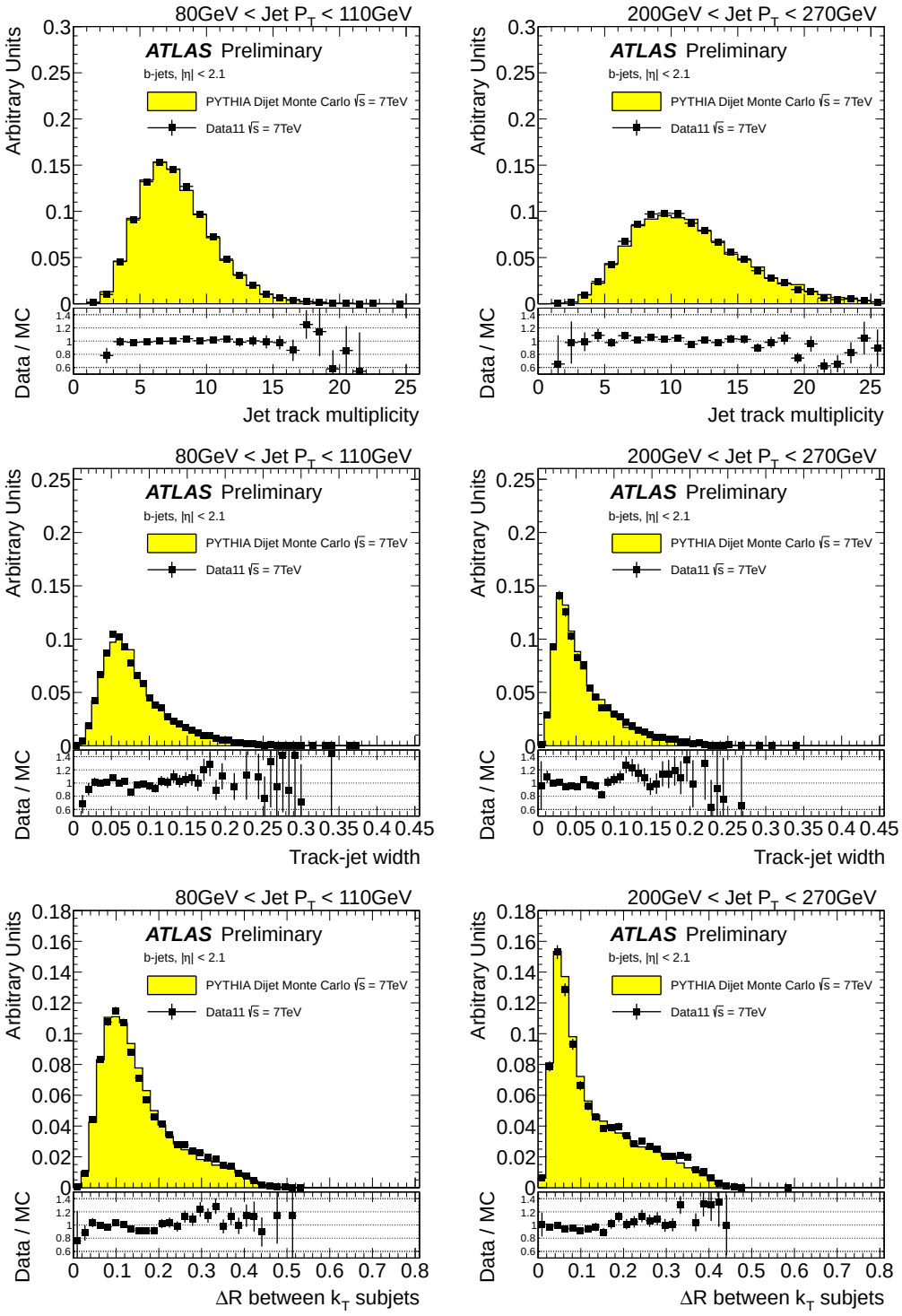

Figure 1.18: Distribution of three tracking variables in 2 different jet p_T bins, for experimental data collected by ATLAS during 2011 (solid black points), and simulated data (filled histograms). The ratio data over simulation is shown at the bottom of each plot.

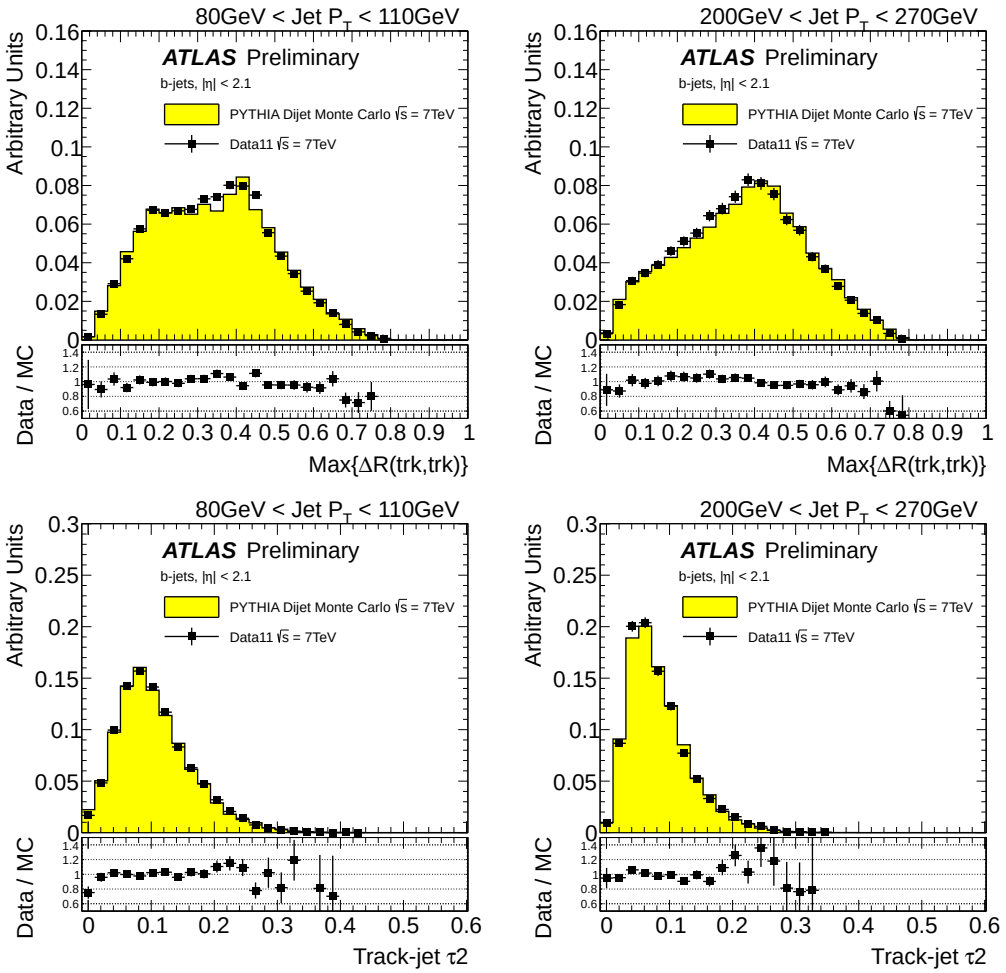

Figure 1.19: Distribution of two tracking variables in two different jet p_T bins, for experimental data collected by ATLAS during 2011 (solid black points), and simulated data (filled histograms). The ratio data over simulation is shown at the bottom of each plot.

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] G. Aad et al. The atlas simulation infrastructure. *The European Physical Journal C*, 70:823–874, 2010.

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

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

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

[7] D. Buskulic et al. Quark and gluon jet properties in symmetric three-jet events. *Physics Letters B*, 384(1-4):353–364, 1996.

[8] 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.

[9] ATLAS Collaboration. Measurement of Jet Mass and Substructure for Inclusive Jetsin $\sqrt{s} = 7$ TeV pp Collisions with the ATLAS Experiment. *ATLAS-CONF-2011-073*, May 2011.

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

[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] G.P. Salam M. Cacciari and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 0804:42, 2008.

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

[17] 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.

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