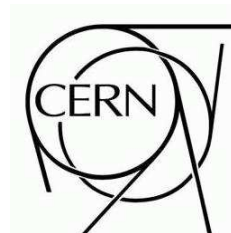




ATLAS NOTE

ATL-PHYS-PUB-2009-000

April 26, 2009



Performance of the ATLAS b -tagging Algorithms

The ATLAS Collaboration¹⁾

This note is part of CERN-OPEN-2008-020. This version of the note should not be cited: all citations should be to CERN-OPEN-2008-020.

Abstract

The ability to identify jets stemming from the fragmentation and hadronization of b quarks is important for the high- p_T physics program of ATLAS: top physics, Higgs boson searches and studies, new phenomena. After an overview of the reconstruction of the key ingredients for b -tagging, the tagging techniques are described. The performance of b -tagging algorithms is then detailed, as well as the impact on performance of several factors and new promising directions. Finally, expected performance in the first data and the anticipated uncertainty with which it can be measured are briefly discussed.

¹⁾This note prepared by G. Aad, J. Bastos, J. Beringer, A. Cheplakov, C. Gwilliam, V. Jain, B. King, C. Lapoire, M. Lehmacher, A. Mehta, A. Rozanov, A. S. Thompson, L. Vacavant, J.-B. de Vivie, M. Vos, and R. Zaidan.

1 Introduction

This note discusses the identification of jets stemming from the hadronization of b quarks, or b -tagging. The ability to identify jets containing b -hadrons is important for the high- p_T physics program of a general-purpose experiment at the LHC such as ATLAS. This is in particular useful to select very pure top samples, to search and/or study Standard Model or supersymmetric (SUSY) Higgs bosons which couple preferably to heavy objects or are produced in association with heavy quarks, to veto the large dominant $t\bar{t}$ background for several physics channels and finally to search for new physics: SUSY decay chains, heavy gauge bosons, etc.

The large majority of these studies requires good b -tagging performance for jets with a transverse momentum ranging from 20 to 150 GeV. However, for super-symmetric processes, jets of p_T as high as 500 GeV may have to be tagged [1], and for exotic phenomena b -jets of up to a few TeV can be produced. For top studies, the signal rates are very high at the LHC and therefore a moderate b -tagging efficiency ($> 50\%$) is acceptable, while a fraction of light jets mis-identified as b -jets below a few per mille suppresses most of the W +jets background (see for instance Ref. [2]). One of the most demanding channels for b -tagging is the production of a light Standard Model Higgs boson in association with a top-antitop pair [3]: $t\bar{t}H(H \rightarrow b\bar{b})$. Four b -jets have to be tagged with very high efficiency ($\epsilon_b \approx 70\%$) since the signal cross-section is low, and the mis-tagging rate must be kept below 1% to fight the large $t\bar{t}$ +jets background.

The identification of b -jets takes advantage of several of their properties which allow us to distinguish them from jets which contain only lighter quarks. First the fragmentation is hard and the b -hadron retains about 70% of the original b quark momentum. In addition, the mass of b -hadrons is relatively high (> 5 GeV). Thus, their decay products may have a large transverse momentum with respect to the jet axis and the opening angle of the decay products is large enough to allow separation. The third and most important property is the relatively long lifetime of hadrons containing a b quark, of the order of 1.5 ps ($c\tau \approx 450\mu\text{m}$). A b -hadron in a jet with $p_T = 50$ GeV will therefore have a significant flight path length $\langle l \rangle = \beta\gamma c\tau$, traveling on average about 3 mm in the transverse plane before decaying. Such displaced vertices can first be identified inclusively by measuring the impact parameters of the tracks from the b -hadron decay products. The transverse impact parameter, d_0 , is the distance of closest approach of the track to the primary vertex point, in the $r - \phi$ projection. The longitudinal impact parameter, z_0 , is the z coordinate of the track at the point of closest approach in $r - \phi$. The tracks from b -hadron decay products tend to have rather large impact parameters which can be distinguished from tracks stemming from the primary vertex. The other more demanding option is to reconstruct explicitly the displaced vertices. These two approaches of using the impact parameters of tracks or reconstructing the secondary vertex will be referred to later on as spatial b -tagging. Finally, the semi-leptonic decays of b -hadrons can be used by tagging the lepton in the jet. In addition, thanks to the hard fragmentation and high mass of b -hadrons, the lepton will have a relatively large transverse momentum and also a large momentum relative to the jet axis. This is the so-called soft lepton tagging (the lepton being soft compared to high- p_T leptons from W or Z decays).

The tagging methods relying on the impact parameter of tracks are detailed in this note. Only a summary and the main results of the other methods are given. The techniques employed to reconstruct either a single inclusive vertex or to attempt to resolve the complex topologies with a secondary b -hadron vertex and a tertiary c -hadron vertex are discussed in Ref. [4], as well as the reconstruction of the primary vertex. The tagging with soft muons or electrons from b -hadron decays is detailed respectively in Ref. [5] and Ref. [6]. The expected performance of the b -tagging algorithms in ATLAS, and the impact of several factors, are explained in detail in this note. However, the assessment of the impact of residual misalignments on the performance is just starting and first results are available in Ref. [7]. While a large effort is put into having a very accurate Monte Carlo simulation, the b -tagging performance must be

measured in data. Several studies aiming at measuring the b -tagging efficiency in dijet events (Ref. [8]) or in $t\bar{t}$ events (Ref. [9]) have been performed. The studies to measure the mis-tagging rates are just starting and are not discussed. Finally, the high-level trigger of ATLAS has the capability to select b -jets. This is particularly interesting for channels with several b -jets where jet thresholds can be lowered at the first level thanks to the b -tagging applied at the second and event-filter levels. The high-level trigger b -tagging performance and strategies are discussed in Ref. [10].

The layout of this note is as follows: in Section 2, the reconstruction of the key objects for b -tagging is briefly explained and the performance summarized. Since the definition of the flavour of a jet is not unambiguous in Monte Carlo, the estimators used to assess the performance are defined in Section 3. Section 4 is intended to be a pedagogical approach to the various tagging algorithms available and to the likelihood ratio formalism used by ATLAS. The b -tagging performance for various physics processes is described in Section 5, relying on the current state-of-the-art b -tagging production software. In Section 6, a few additional studies aiming at better understanding some critical aspects of the b -tagging are detailed, while in Section 7 three studies showing new directions to improve the b -tagging performance are presented. In both cases the studies are described in a separate section because either they required specific datasets or they relied on software and/or cuts/optimizations which were different from the ones currently in use in the ATLAS software or they even required new software developments. In addition, the anticipated uncertainty with which the b -tagging performance may be measured in data is discussed in Section 8. Finally in Section 9, the main findings and the expected performance in the first data are summarized.

2 Reconstruction of the key objects

The reconstruction of the various objects needed for b -tagging and its performance are summarized in this section.

2.1 Charged tracks

The tracks reconstructed in the ATLAS Inner Detector [11] are the main ingredient for b -tagging. On average a track consists of 3 pixel hits, 4 space-points in the silicon micro-strip detector and about 36 hits in the Transition Radiation Tracker (TRT). The innermost pixel layer (the so-called b -layer) is located at a radius of 5 cm, while the TRT extends up to a radius of 1 m. The tracker is immersed in a 2 T magnetic field generated by the central solenoid. The intrinsic measurement accuracy of the pixels is around $10\ \mu\text{m}$ in $r\phi$ and $115\ \mu\text{m}$ in z . All these allow the tracker to measure efficiently and with good accuracy the tracks within $|\eta| < 2.5$ and down to $p_T \sim 500\ \text{MeV}$. For a central track with $p_T = 5\ \text{GeV}$, which is typical for b -tagging, the relative transverse momentum resolution is around 1.5% and the transverse impact parameter resolution is about $35\ \mu\text{m}$. Further details can be found in Refs. [11, 12].

Most of the results in this note are based on the default pattern-recognition and fitting algorithm, NewTracking. Its performance is described in Ref. [12]. When relevant, some comparisons are made with an alternate algorithm, iPatRec.

2.1.1 Baseline track selection

The track selection for b -tagging is designed to select well-measured tracks and reject fake tracks and tracks from long-lived particles (K_s, Λ or other hyperon decays) and material interactions (photon conversions or hadronic interactions).

Two different quality levels are used. For the standard quality level, at least seven precision hits (pixel or micro-strip hits) are required. The transverse and longitudinal impact parameters at the perigee

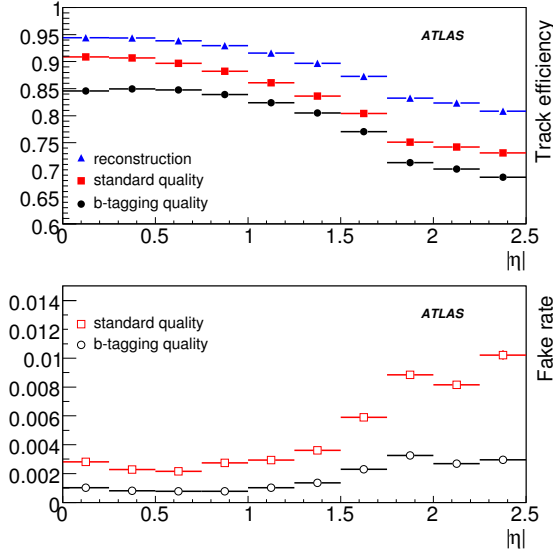


Figure 1: Tracking efficiency (top plot) and fake rate (bottom plot) versus track pseudo-rapidity, for three levels of track selection: matching (blue triangles), standard quality cuts (red squares) and b -tagging quality cuts (black circles), in $t\bar{t}$ events.

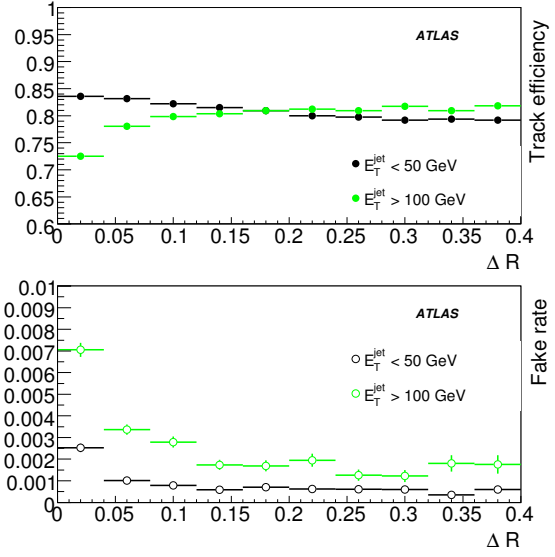


Figure 2: Tracking efficiency (top plot) and fake rate (bottom) versus distance to jet axis, for tracks fulfilling the b -tagging quality cuts and associated to low- p_T jets (black symbols) or high- p_T jets (green symbols), in $t\bar{t}$ events.

must fulfil $|d_0| < 2$ mm and $|z_0 - z_{pv}| \sin \theta < 10$ mm respectively, where z_{pv} is the longitudinal location of the primary vertex. Only tracks with $p_T > 1$ GeV are considered. For the b -tagging quality, the extra requirements are: at least two hits in the pixel detector of which one must be in the b -layer, as well as $|d_0| < 1$ mm and $|z_0 - z_{pv}| \sin \theta < 1.5$ mm. This selection is used by all the tagging algorithms relying on the impact parameters of tracks, while slightly different selections are used by the secondary vertex algorithms as discussed in Ref. [4].

2.1.2 Tracking efficiency

The b -tagging performance strongly depends upon the tracking efficiency. The tracking performance inside jets, where the track density may be high, is discussed in the following. The tracking performance for single tracks is discussed in Ref. [12].

Figure 1 shows the tracking efficiency and fake rate for tracks in $t\bar{t}$ events as a function of the track pseudo-rapidity. For the efficiency denominator, only charged primary pions²⁾ produced well before the b -layer ($|x - x_{pv}| < 10$ mm, $|y - y_{pv}| < 10$ mm) and with $p_T > 1$ GeV and $|\eta| < 2.5$ are considered. The first level of the efficiency corresponds to the basic reconstruction efficiency, where a track matched to a Monte Carlo particle is found. The fake rate is defined as the fraction of reconstructed tracks which do not pass the matching criteria used for the efficiency, i.e. less than 80% of their hits are coming from the same Monte Carlo particle. At high pseudo-rapidities, the tracking performance deteriorates mostly because of increased material and more ambiguous measurements.

Figure 2 shows the tracking efficiency and fake rate for tracks in $t\bar{t}$ events as a function of their distance $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ to the axis of the closest jet, for tracks fulfilling the b -tagging quality cuts. The tracking performance degrades near the core of the jet where the track density is the highest and induces pattern-recognition problems. This is especially visible for high- p_T (> 100 GeV) jets.

²⁾For the tracking studies only pions were considered, but similar results are expected for charged kaons, protons, etc. which are all used for b -tagging.

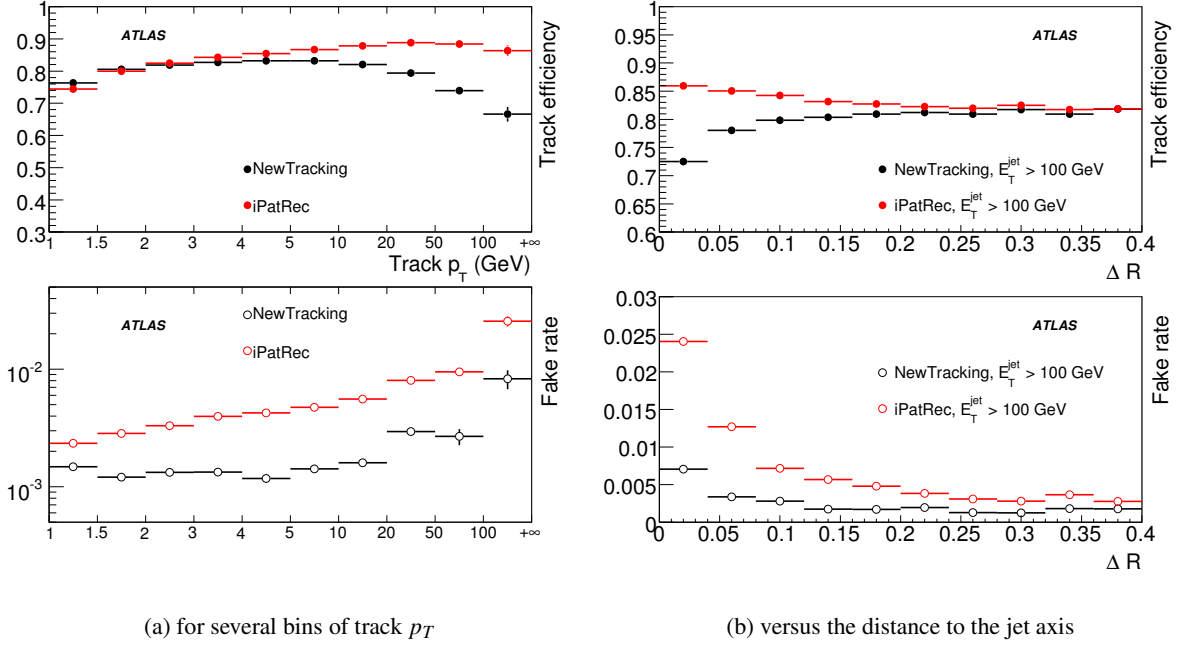


Figure 3: Tracking efficiency (top plots) and fake rate (bottom plots) in $t\bar{t}$ events after the b -tagging quality cuts, for two tracking algorithms: default NewTracking (black symbols) and iPatRec (red symbols).

Finally, in Figure 3 the tracking efficiency and fake rates obtained with the default algorithm and with iPatRec are compared. The first plot shows the comparison for several bins in the track p_T for all jets, while the second plot is as a function of the distance to the jet axis for jets with $E_T > 100$ GeV. It is interesting to note that the two algorithms have a different working point: the default algorithm maintains a low level of fakes at the price of losing in efficiency, while the complementary choice was taken for iPatRec. This difference in treatment will lead to different b -tagging performance for jets with high momentum, as discussed in particular in Section 5.6. The features seen in these plots are specific to the pattern-recognition inside jets: for instance the decrease of the NewTracking efficiency at high track p_T is not visible for isolated tracks; it is here correlated with the local track density since high- p_T tracks are more likely to originate from denser high- p_T jets.

2.1.3 Tracks with shared hits

Tracks originating from the same point and passing the track selection cuts will not necessarily have the same impact parameter distributions. First of all, even using the track parameters normalized to their error will not compensate for all resolution effects, such as non-Gaussian tails. In addition, the pattern-recognition process itself can produce tracks of variable quality depending on their hit contents. Those tracks require a special treatment to be flagged appropriately. The most significant subset of such tracks is formed by the tracks which are sharing some of their hits with other tracks.

Figure 4 shows the fraction of tracks which are sharing at least one hit with another reconstructed track versus the distance of the track to the jet axis, for jets originating from $t\bar{t}$ events. Currently for b -tagging purposes, a track is defined as a track with shared hits if it has at least one shared hit in the pixels or two shared hits in the strips. As expected, the fraction of tracks with shared hits increases with the local track density, and is therefore higher for high- p_T jets and in the core of the jets. In $t\bar{t}$ events

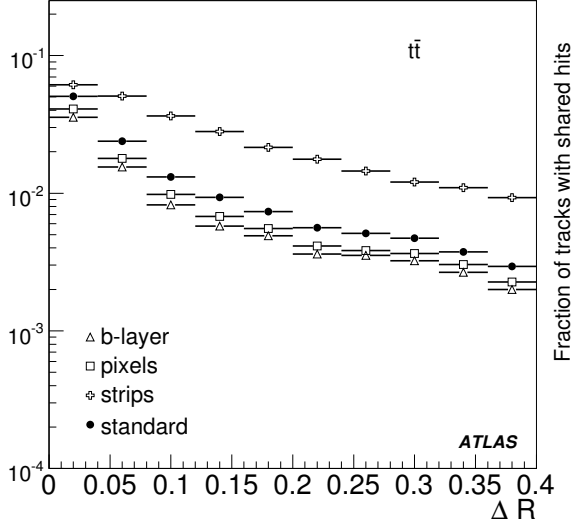


Figure 4: Fraction of tracks with shared hits versus distance to the jet axis. Tracks fulfilling the b -tagging quality cuts, and with at least one shared hit in the silicon systems are shown. The standard definition of shared hits (see text) is shown as well.

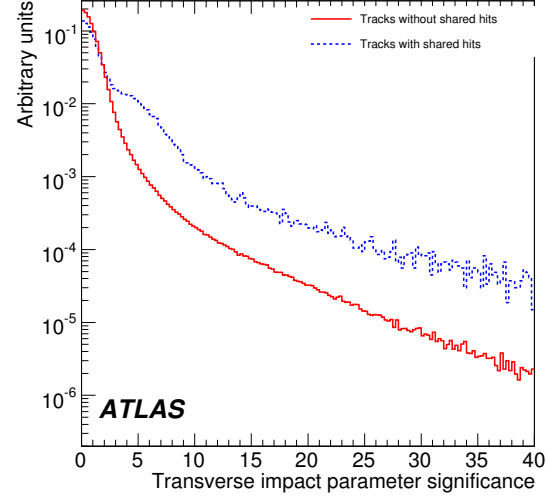


Figure 5: Transverse impact parameter significance d_0/σ_{d_0} for tracks in light jets. Two categories of tracks are used: regular ones (red plain curve) and tracks with shared hits (blue dashed). Both distributions are normalized to unity.

the average p_T for taggable (*i.e.* $p_T > 15$ GeV and $|\eta| < 2.5$) b -jets and light jets are respectively 74 and 55 GeV. The fraction of tracks with shared hits is about 2%. For jets with a transverse momentum of about 140 GeV (WH events with $m_H=400$ GeV, see below), this fraction is twice as high. In both cases the fraction is roughly similar for NewTracking and iPatRec. In an extreme case, for $Z' \rightarrow b\bar{b}$ events with $m_{Z'} = 2$ TeV, a majority of tracks have shared hits and the fraction depends significantly on the reconstruction algorithm (cf. Section 5.6). Even when the overall level of shared hits is relatively low, it has been demonstrated that those tracks should be treated appropriately since their impact on the b -tagging performance is significant. Indeed, the impact parameter significances, defined as the ratios d_0/σ_{d_0} and z_0/σ_{z_0} of the impact parameters to their measured error, for tracks in light jets exhibit a very different behavior depending on whether the track is a regular one or a track with shared hits, as shown in Figure 5. It is clear that tracks with shared hits can mimic lifetime more easily.

2.1.4 Impact parameter resolution

The resolution of the track impact parameter is a crucial ingredient to be able to discriminate tracks coming from long-lived hadrons and prompt tracks. To estimate it, all the reconstructed tracks in $t\bar{t}$ events fulfilling the b -tagging quality cuts and matched to a good Monte-Carlo track as defined in Section 2.1.2 were used. The difference between the reconstructed and the true impact parameter within a bin was fitted with a single gaussian, whose σ is reported on Figures 6(a) and 6(b), for respectively the transverse and longitudinal impact parameters. For a central track with $p_T = 5$ GeV, which is typical for b -tagging, the transverse impact parameter resolution is about $35 \mu\text{m}$.

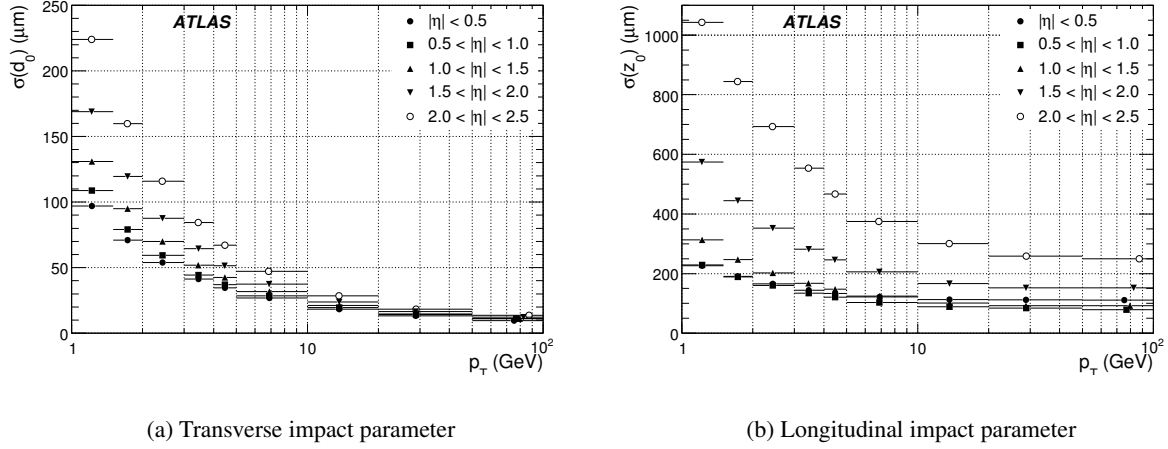


Figure 6: Track impact parameter resolution versus track p_T , for several bins in the track pseudo-rapidity.

2.2 Primary vertex finding

Another key ingredient for b -tagging is the primary vertex of the event. The impact parameters of tracks are recomputed with respect to its position and tracks compatible with the primary vertex are excluded from the secondary vertex searches. At LHC the beam-spot size will be $\sigma_{xy} = 15 \mu\text{m}$ and $\sigma_z = 5.6 \text{ cm}$: therefore the primary vertex is especially important for the z direction, while in the transverse plane only the beam-line could be used. The strategies to find the primary vertex and their performance are explained in Ref. [4]. The efficiency to find the primary vertex is very high in the high- p_T events of interest, and the resolution on its position is around $12 \mu\text{m}$ in each transverse direction and $50 \mu\text{m}$ along z . With pile-up, the presence of additional minimum bias vertices makes the choice of the primary vertex less trivial: at a luminosity of $2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ (on average 4.6 minimum bias events per bunch-crossing) a wrong vertex can be picked up as the primary vertex in about 10% of the cases [4], thus causing a deterioration in the b -jet tagging efficiency.

2.3 Jet algorithms

The baseline jet algorithm for the studies in this note is a seeded cone algorithm using the calorimeter towers with a cone size of $\Delta R = 0.4$, and where the cells were calibrated using the H1 method (see Ref. [13] for details). The impact on b -tagging performance of using other jet algorithms is discussed in Section 6.4.

For b -tagging purposes, only the jet direction is relevant. In the first place, this direction is used to define which tracks should be associated with the jets. The actual tagging is done on this subset of tracks. Currently tracks within a distance $\Delta R < 0.4$ of the jet axis are associated to the jet. A given track is associated to only one jet (the closest in ΔR). This is the case for actually any jet collections, regardless of the cone size of the jet. The jet direction is also used to sign the impact parameters of the tracks in the jet as explained in Section 4.1.2.

Except when stated otherwise, there was no attempt to remove from the reconstructed jet collection the jets which are composed of only electrons. In the $t\bar{t}$ sample (semi-leptonic and di-leptonic channels), about 5% of the taggable reconstructed jets are electrons. There is no dedicated treatment for muons. Isolated muons are very unlikely to fake jets, at least for the common processes under consideration in this note where $p_T(\mu) < 100 \text{ GeV}$. Muons in jets, stemming from b/c -hadron semi-leptonic decays

and measured in the muon spectrometer, deposit on average about 3 GeV in the calorimeter but their momentum as measured in the inner detector and the muon system is not used to refine the kinematics of the jet, which remain purely calorimeter-based.

Only jets fulfilling $p_T > 15$ GeV and $|\eta| < 2.5$ are deemed taggable and considered in the performance studies.

2.4 Soft lepton reconstruction

Leptons arising from semi-leptonic decays of b -hadrons or subsequent c -hadrons can be used to tag b -jets.

Soft muons are reconstructed [5] using two complementary reconstruction algorithms. A *combined muon* corresponds to a track fully reconstructed in the muon spectrometer that matches a track in the inner detector. Low-momentum muons (below $p \sim 5$ GeV) which cannot reach the muon middle and outer stations are identified by matching an inner detector track with a segment in the muon spectrometer inner stations. Muons satisfying some basic requirements ($p_T > 3$ GeV, $|d_0| < 4$ mm) are associated to the closest jet provided that $\Delta R < 0.5$. Finally, the kinematic properties of the jet-muon system are used in order to reject the background caused by punch-throughs and decays-in-flight in light jets.

Reconstructing soft electrons [6] in the calorimeter inside a jet is more difficult. This is achieved by matching an inner detector track to an electromagnetic cluster. For a given track, only the energy contained in a small window around the track extrapolation is used. The contribution of neighbouring hadronic showers is therefore reduced. The identification procedure takes full advantage of the tracking capabilities of the inner detector as well as of the granularity of the electromagnetic calorimeter: a likelihood ratio combines inner detector information such as transition radiation hits with shower shape variables from the calorimeter. The performance is, however, highly dependent on the track density in jets as well as the quantity of matter in front of the electromagnetic calorimeter.

3 Performance estimators

3.1 Labelling

To define b -tagging performance, the Monte Carlo event history is used to know the type of parton from which a jet originates. This *labelling* procedure is not unambiguous and is not strictly identical for different Monte Carlo generators. For the results presented here, a quark labelling has been used: a jet is labelled as a b -jet if a b quark with $p_T > 5$ GeV is found in a cone of size $\Delta R = 0.3$ around the jet direction. The various labelling hypotheses are tried in this order: b quark, c quark and τ lepton. When no heavy flavour quark nor τ lepton satisfies these requirements, the jet is labelled as a light-jet. No attempt is made to distinguish between u , d , s quarks and gluon since such a label is even more ambiguous.

3.2 Efficiency and rejection

For performance studies, only jets fulfilling $p_T > 15$ GeV and $|\eta| < 2.5$ are considered and referred to as taggable jets. In the following, jets for which no track passed the b -tagging quality cuts are still counted in the performance estimators. However, events where the primary vertex could not be reconstructed are ignored. In addition, b -jets were not categorized according to the nature of the b -hadron decay: b -jets with semi-leptonic decays behave quite differently from jets with hadronic decays, even when using purely spatial methods, but in the following no distinction was made.

The tagging efficiency is naturally defined as the fraction of taggable jets labelled as b -jets (see previous section) which are actually tagged as b -jets by the tagging algorithm under study. The mis-

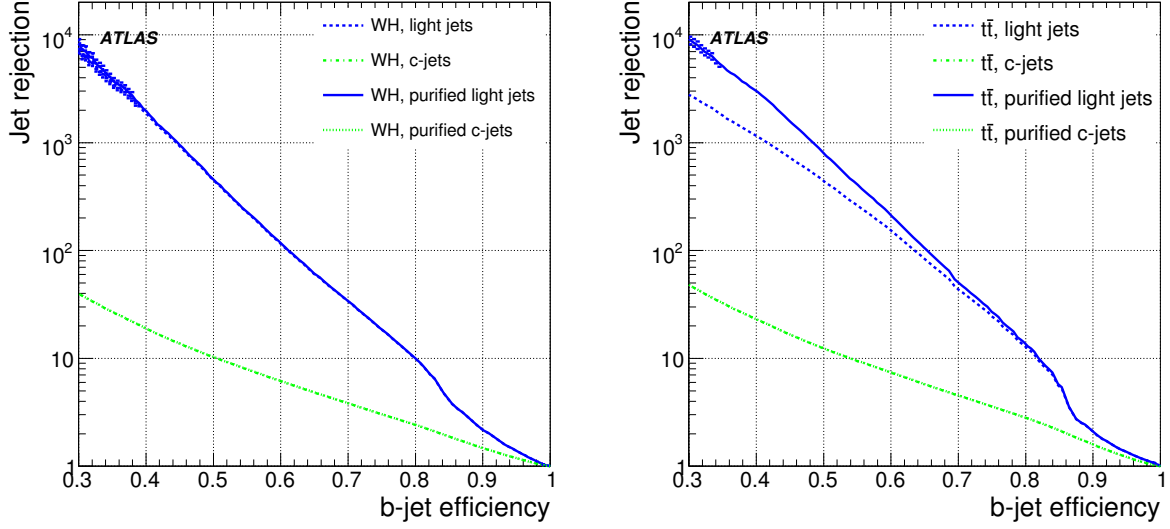


Figure 7: Rejection of light jets and c -jets with and without purification versus b -jet efficiency for WH ($m_H = 120$ GeV) and $t\bar{t}$ events, using the tagging algorithm based on 3D impact parameter and secondary vertex.

tagging rate is the fraction of taggable jets not labelled as b which are actually tagged as b -jets. For historical reasons the jet rejection is used instead: this is simply the inverse of the mis-tagging rate.

3.3 Purification

A difficulty arises as soon as the jet multiplicity is high and various jet flavours are present in a single event: a jet with $\Delta R(\text{jet} - b) = 0.31$ is labelled as a light jet, although tracks from b -hadron decay with high lifetime content are likely to be associated to it.

This leads to a decrease of the estimated performance, not related to the b -tagging algorithm itself but to the labelling procedure which strongly depends on the activity of the event. In order to obtain a more reliable estimation of b -tagging performance, a purification procedure has been devised: light jets for which a b quark, a c quark or a τ lepton is found within a cone of size $\Delta R = 0.8$ around the jet direction are not used to compute the rejection.

The performance estimated after purification represents the intrinsic power of the b -tagging algorithms and should be similar for different kinds of hard event, whereas results obtained for the complete light jet sample are more dependent on the event type. On the other hand, the latter is more representative of the actual b -tagging power for a given physics analysis. This is illustrated in Figure 7: the light jet rejection in simple WH events is similar without or with purification (left plot), while for busier $t\bar{t}$ events (right plot) the two curves differ in the region where lifetime content as opposed to resolution effects dominates (*i.e.* for $\epsilon_b < 80\%$). In the following, jets fulfilling the purification procedure will be referred to as purified or pure jets, the ones failing this procedure will be called non-pure jets, while all the jets will be called raw jets.

4 b -tagging algorithms

In this section the various algorithms used in ATLAS to tag b -jets are explained. The spatial algorithms, built on tracks and subsequently vertices, are the most powerful ones. Most of them are based on a

likelihood ratio approach, but simpler and more robust tagging algorithms are also available. Soft lepton tagging algorithms are also very important, in particular since the correlation with the previous ones is minimal. Their performance is summarized in section 4.3.

4.1 Spatial algorithms based on likelihood ratio

All tracks in the jet fulfilling the b -tagging quality cuts described in 2.1.1 are considered for the spatial b -tagging algorithms. In typical $t\bar{t}$ events, the average number of those tracks per light (b -) jet is 3.7 (5.5) and their average p_T is 6.6 (6.3) GeV, respectively.

4.1.1 V^0 and secondary interactions rejection

The preselection cuts on impact parameters reject a large fraction of long-lived particles and secondary interactions. Among the remaining tracks, the ones identified by the secondary vertex search (section 4.1.3) as likely to come from V^0 decays are rejected (they amount to between 1% and 3% of the tracks in light and b -jets respectively). To do so, the search starts by building all two-track pairs that form a good vertex. The mass of the vertex is used to reject the tracks which are likely to come from K_s, Λ decays and photon conversions. The radius of the vertex is compared to a crude description of the innermost pixel layers to reject secondary interactions in material. The cuts and performance of this selection are described in Ref. [4].

4.1.2 Impact parameter tagging algorithms

For the tagging itself, the impact parameters of tracks are computed with respect to the primary vertex (cf. section 2.2). On the basis that the decay point of the b -hadron must lie along its flight path, the impact parameter is signed to further discriminate the tracks from b -hadron decay from tracks originating from the primary vertex. The sign is defined using the jet direction \vec{P}_j as measured by the calorimeters (cf. section 2.3), the direction \vec{P}_t and the position \vec{X}_t of the track at the point of closest approach to the primary vertex and the position \vec{X}_{pv} of the primary vertex:

$$\text{sign}(d_0) = (\vec{P}_j \times \vec{P}_t) \cdot \left(\vec{P}_t \times (\vec{X}_{pv} - \vec{X}_t) \right)$$

The experimental resolution generates a random sign for the tracks originating from the primary vertex, while tracks from the b/c hadron decay tend to have a positive sign. The sign of the longitudinal impact parameter z_0 is given by the sign of $(\eta_j - \eta_t) \times z_{0t}$ where again the t subscript refers to quantities defined at the point of closest approach to the primary vertex.

The distribution of the signed transverse impact parameter d_0 is shown on Figure 8, left plot, for tracks coming from b -jets, c -jets and light jets. The right plot shows the significance distribution d_0/σ_{d_0} which gives more weight to precisely measured tracks. Combining the impact parameter significances of all the tracks in the jet is the basis of the first method to tag b -jets. Three tagging algorithms are defined in this way: IP1D relies on the longitudinal impact parameter, IP2D on the transverse impact parameter and finally IP3D which uses two-dimensional histograms of the longitudinal versus transverse impact parameters, taking advantage of their correlations.

4.1.3 Secondary vertex tagging algorithms

To further increase the discrimination between b -jets and light jets, the inclusive vertex formed by the decay products of the bottom hadron, including the products of the eventual subsequent charm hadron decay, can be sought. The reader is referred to Ref. [4] for all details. The search starts by building all two-track pairs that form a good vertex, using only tracks far enough from the primary vertex

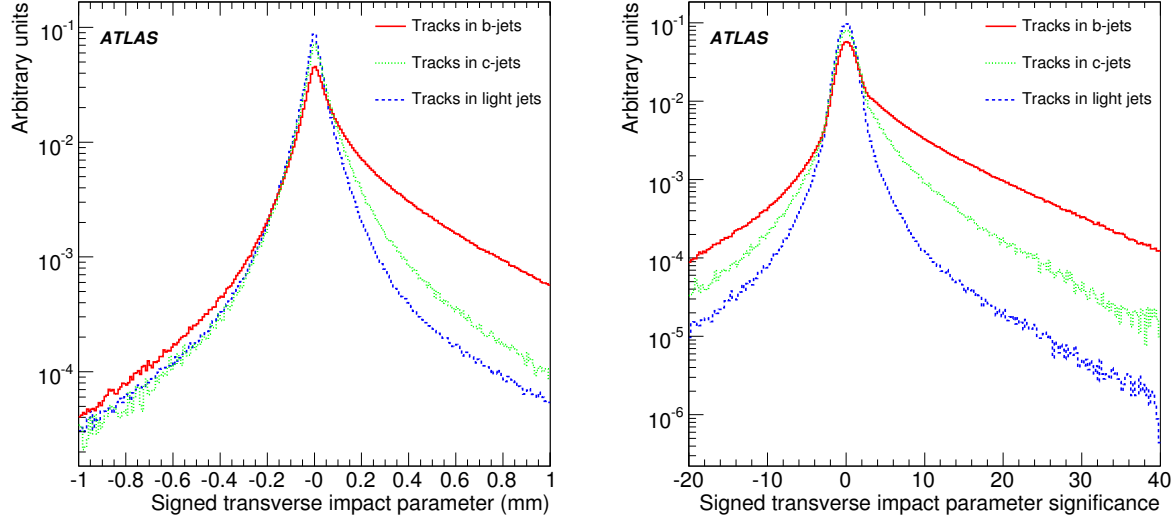


Figure 8: Signed transverse impact parameter d_0 distribution (left) and signed transverse impact parameter significance d_0/σ_{d_0} distribution (right) for b -jets, c -jets and light jets.

($L_{3D}/\sigma_{L_{3D}} > 2$ where $L_{3D} \equiv \|\vec{X}_{pv} - \vec{X}_t\|$ is the three dimensional distance between the primary vertex and the point of closest approach of the track to this vertex). Vertices compatible with a V^0 or material interaction are rejected. All tracks from the remaining two-track vertices are combined into a single inclusive vertex, using an iterative procedure to remove the worst track until the χ^2 of the vertex fit is good. Three of the vertex properties are exploited: the invariant mass of all tracks associated to the vertex, the ratio of the sum of the energies of the tracks participating to the vertex to the sum of the energies of all tracks in the jet and the number of two-track vertices. These properties are illustrated in Figure 9 for b -jets and light jets. The so-called SV tagging algorithms make different use of these properties: SV1 relies on a 2D-distribution of the two first variables and a 1D-distribution of the number of two-track vertices, while SV2 is based on a 3D-histogram of the three properties which requires quite some statistics. The secondary vertex finding efficiency depends in particular on the event topology, but the typical efficiency ε_b^{SV} is higher than 60% in b -jets. The SV taggers require an a priori knowledge of ε_b^{SV} and ε_u^{SV} .

A completely new algorithm, JetFitter, is also available, which exploits the topological structure of weak b - and c -hadron decays inside the jet. A Kalman filter is used to find a common line on which the primary vertex and the beauty and charm vertices lie, as well as their position on this line approximating the b -hadron flight path. With this approach, the b - and c -hadron vertices are not merged, even when only a single track is attached to each of them. The discrimination between b -, c - and light jets is based on a likelihood using similar variables to the SV tagging algorithm above, and additional variables such as the flight length significances of the vertices. This algorithm and its performance are also described in detail in Ref. [4].

4.1.4 Formalism of likelihood ratio

For both the impact parameter tagging and the secondary vertex tagging, a likelihood ratio method is used: the measured value S_i of a discriminating variable is compared to pre-defined smoothed and normalized distributions for both the b - and light jet hypotheses, $b(S_i)$ and $u(S_i)$. Two- and three-dimensional probability density functions are used as well for some tagging algorithms. The ratio of the probabilities $b(S_i)/u(S_i)$ defines the track or vertex weight, which can be combined into a jet weight W_{Jet} as the sum

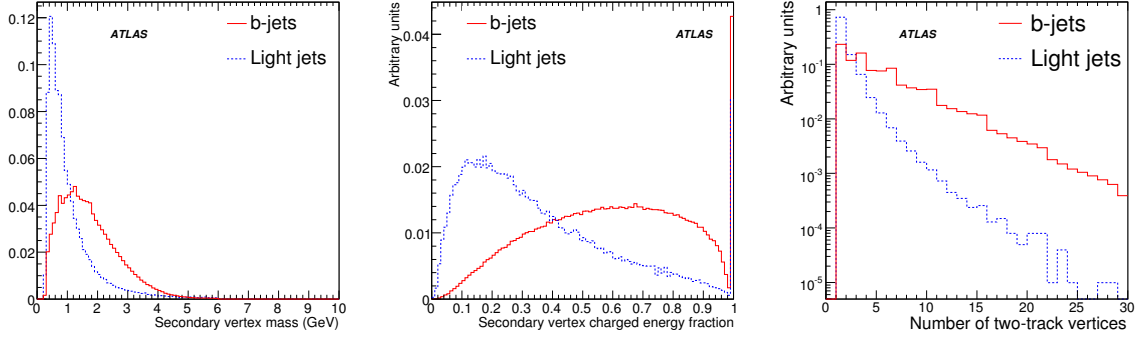


Figure 9: Secondary vertex variables: invariant mass of all tracks in vertex (left), energy fraction vertex/jet (center) and number of two-track vertices (right) for b -jets and light jets.

of the logarithms of the N_T individual track weights W_i :

$$W_{Jet} = \sum_{i=1}^{N_T} \ln W_i = \sum_{i=1}^{N_T} \ln \frac{b(S_i)}{u(S_i)} \quad (1)$$

The distribution of such a weight is shown in Figure 10 for b -, c - and light jets for two different tagging algorithms: IP2D and the sum of the weights from IP3D and SV1. When no vertex is found, the SV taggers return a weight of $\ln \frac{1-\epsilon_b^{SV}}{1-\epsilon_u^{SV}}$. To select b -jets, a cut value on W_{Jet} must be chosen, corresponding to a given efficiency. The relation between the cut value and the efficiency depends on the jet transverse momentum and rapidity, and therefore is different for different samples.

4.1.5 Likelihood ratio and track categories

As seen already, tracks may exhibit different behavior even after the track selection, such as the tracks with shared hits (Figure 4). One idea to take advantage of the different properties of tracks is to arrange all tracks into various categories and use dedicated probability density functions for each category. The likelihood ratio formalism permits to incorporate such categories in a straightforward way. After the division of the tracks into disjoint categories j , where every category has its own set of reference histograms b_j and u_j , the jet weight can simply be written as the sum over all tracks in each category N_T^j and all categories N_C :

$$W_{Jet} = \sum_{j=1}^{N_C} \left(\sum_{i=1}^{N_T^j} \ln \frac{b_j(S_i)}{u_j(S_i)} \right) \quad (2)$$

Currently in the b -tagging software, two track categories are used: the *Shared* tracks (tracks with shared hits), and the complementary subset of tracks called *Good* tracks. These track categories are only used for the time being for the IP1D, IP2D and IP3D tagging algorithms.

4.2 Other spatial algorithms

The spatial algorithms based on likelihood ratios require an a-priori knowledge of the properties of both b -jets and light jets. Methods to measure them in data are being devised for the b -jets [8, 9] but will require at least about 100 pb^{-1} . In addition, there is no clear way to extract a pure enough sample of light jets, and Monte Carlo simulation will probably have to be used once a thorough validation against

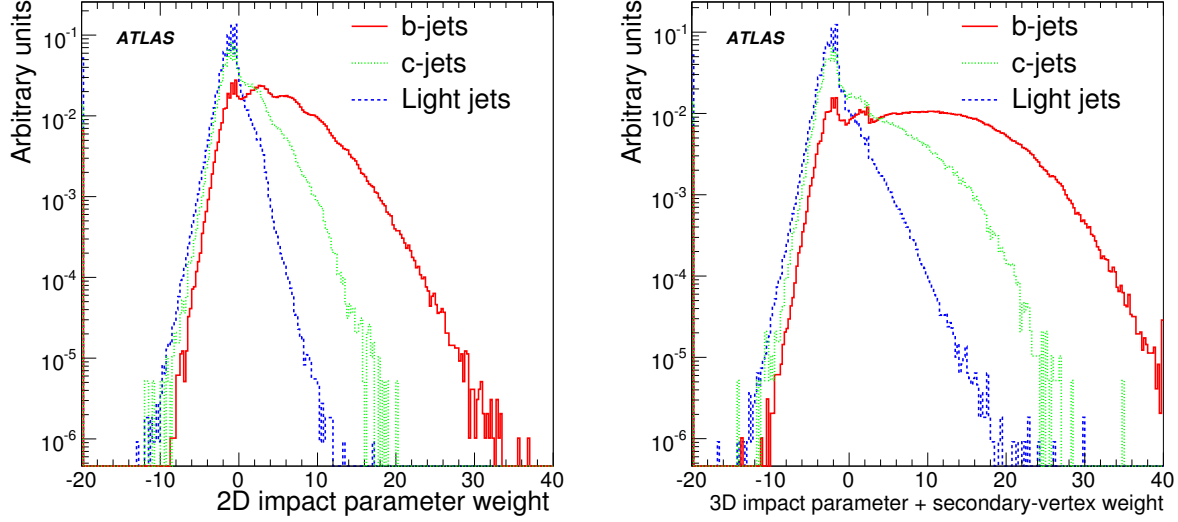


Figure 10: Jet b -tagging weight distribution for b -jets, c -jets and purified light jets. The left plot is for the IP2D tagging algorithm. The right plot corresponds to the IP3D+SV1 tagging algorithm.

data has been performed. A few other spatial tagging algorithms, less powerful, are therefore developed, which have less reliance on Monte Carlo and are expected to be easier and faster to commission with the first real data.

The simplest approach that could be used, at least at the beginning, is the counting of tracks with large impact parameter or large impact parameter significance. Requiring a few of these tracks provides a sample enriched in b -jets. The performance of such a tagging algorithm is not discussed in this note because it is not yet fully implemented in ATLAS. Such a simple tagger may also be very useful at the trigger level.

Another approach is to combine the impact parameter of all the tracks in the jet. JetProb is an implementation of the ALEPH tagging algorithm [14], used extensively at LEP and later at the Tevatron. The signed impact parameter significance d_0/σ_{d_0} of each selected track in the jet is compared to a resolution function \mathcal{R} for prompt tracks, in order to measure the probability that the track i originates from the primary vertex (Figure 11(a)):

$$\mathcal{P}_i = \int_{-\infty}^{-|d_0^i/\sigma_{d_0}^i|} \mathcal{R}(x) dx \quad (3)$$

The resolution function can be measured in data using the negative side of the signed impact parameter distribution (cf. section 6.5.1), assuming there is no contribution from heavy-flavour particles which is not strictly true.

The individual probability of each of the N tracks associated to the jet are then combined to obtain a jet probability \mathcal{P}_{jet} which discriminates b -jets against light jets (Figure 11(b)):

$$\mathcal{P}_{jet} = \mathcal{P}_0 \sum_{j=0}^{N-1} \frac{(-\ln \mathcal{P}_0)^j}{j!} \quad (4)$$

where

$$\mathcal{P}_0 = \prod_{i=1}^N \mathcal{P}'_i \quad \text{and} \quad \begin{cases} \mathcal{P}'_i = \frac{\mathcal{P}_i}{2} & \text{if } d_0^i > 0 \\ \mathcal{P}'_i = \left(1 - \frac{\mathcal{P}_i}{2}\right) & \text{if } d_0^i < 0 \end{cases} \quad (5)$$

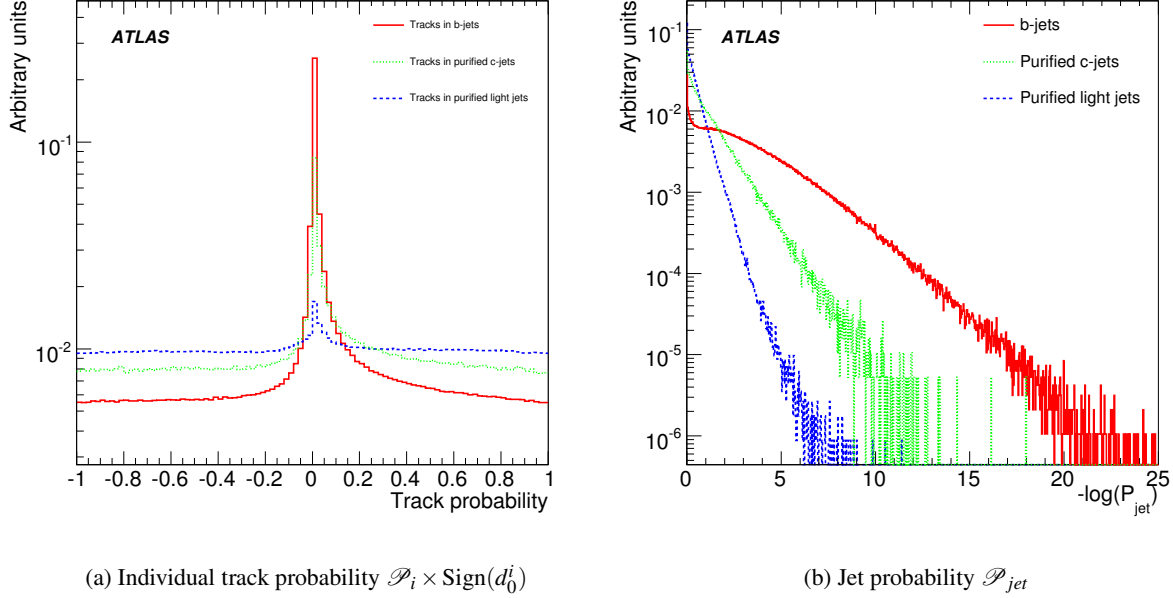


Figure 11: Distributions of the probability of compatibility with the primary vertex for individual tracks (left plot) and for all tracks in the jet (right plot) as defined for JetProb. The cases of b -jets (red plain), c -jets (green dashed) and light jets (blue dotted line) are shown.

4.3 Soft lepton algorithms

Soft lepton tagging relies on the semi-leptonic decays of bottom and charm hadrons. Therefore it is intrinsically limited by the branching ratios to leptons: at most 21% [15] of b -jets will contain a soft lepton of a given flavour, including cascade decays of bottom to charm hadrons. However, tagging algorithms based on soft leptons exhibit very high purity and low correlations with the track-based tagging algorithms, which is very important for checking and cross-calibrating performance in data (see for instance Ref. [8]).

4.3.1 Soft muons

Once a reconstructed muon is associated to a jet as explained briefly in Section 2.4, a likelihood permits to discriminate light jets from b -jets. The algorithm and its performance are detailed in Ref. [5] and will not be discussed further in this note. To summarize, a light jet rejection of about 300 can be achieved for a b -tagging efficiency of 10%. Those numbers include the semi-leptonic branching ratio, the detector acceptance, the reconstruction efficiency as well as the jet-muon association efficiency. This was estimated in $t\bar{t}$ events, including a simulation of the cavern background (low-energy neutrons and photons stemming from the interaction of forward particles with the shielding) which reduces the rejection level by about 30%. The performance is relatively steady in the jet p_T range 15-100 GeV and in pseudo-rapidity.

4.3.2 Soft electrons

A likelihood ratio is also used for soft electrons. The algorithm and its performance are detailed in Ref. [6] and will not be discussed further in this note. A light jet rejection of about 100 can be achieved for a b -tagging efficiency of 7%. The efficiency of the soft electron identification is high, since two-thirds

of the true b -jets containing a real soft electron are tagged by the soft electron algorithm. However, about 25% of light jets are mis-tagged by real electrons from photon conversions and Dalitz decays. This was estimated in WH ($m_H = 120$ GeV) events without pile-up. Based on previous study [16], a further degradation by 10% (30%) is expected when on average 4.6 (23) minimum-bias events are added. While the performance is constant in jet p_T in the range 15-100 GeV, it degrades quickly with the jet pseudo-rapidity because of the higher amount of dead material, the poor performance in the transition region between the barrel and end-cap cryostats of the electromagnetic calorimeter ($1.37 < |\eta| < 1.52$) and the absence of the TRT beyond $|\eta| > 2$.

4.4 Combining tagging algorithms

Currently only the likelihood-based tagging algorithms have been combined, since the formalism is easy in this case: the weights of the individual tagging algorithms are simply summed up. The most commonly used tagging algorithm, IP3D+SV1, is actually such a combination. It should be noted that the SV tagging algorithms have been optimized to work in conjunction with the IP ones. Another one combines IP3D and JetFitter. Multivariate approaches to combine all tagging algorithms, including the soft lepton ones, have not received much attention so far. There are, however, some new studies and the use for instance of boosted decision trees is discussed in Section 7.3.

4.5 Calibration of tagging algorithms

The likelihood-based tagging algorithms require knowledge of the probability density functions of the discriminating variables for both the b - and light jet hypotheses: this is called the calibration of the tagging algorithms, or their reference histograms. In the following, those functions have been derived from a large sample of jets coming from $t\bar{t}$ and $t\bar{t}jj$ events. Several issues about the calibration and its impact on b -tagging performance are discussed in Section 6.5.

5 Performance for various physics processes

In this section, the b -tagging performance is reviewed for several physics channels of interest. Several spatial tagging algorithms are considered: JetProb and IP2D which are best suited for the initial period, IP3D and then IP3D+SV1 for regular operations once the secondary vertexing is understood and finally IP3D+JetFitter for the ultimate performance.

5.1 Dependence on jet transverse momentum and pseudo-rapidity

The spatial b -tagging performance depends strongly on the jet momentum and rapidity: the p_T and η dependencies of the b -tagging efficiency and light jet rejection for a given cut on the b -tagging weight are shown in Figure 12. At high p_T or at high $|\eta|$, the b -jet tagging performance is poor, regardless of which tagging algorithm is used. At low p_T , maintaining a reasonable b -jet efficiency is possible only by loosening the cut on the weight, at the price of a very low rejection of light jets. The strong dependence, especially in p_T , makes the extraction of the b -jet efficiency from data complicated and means that more integrated luminosity will be required, since several bins are needed.

Because of these strong p_T and η dependencies, and since various samples have very different spectra, it is not straightforward to compare between channels the integrated rejection numbers shown in the following. It is worth noting that this dependence is really a two-dimensional one, thus Figure 12 is useful for illustrative purposes but the $(p_T, |\eta|)$ spectrum of the jets in the sample considered is not properly factorized out. This is important for instance to parametrize the b -tagging performance. This is further illustrated in Figure 13: the rejections achieved in two different samples become more similar as

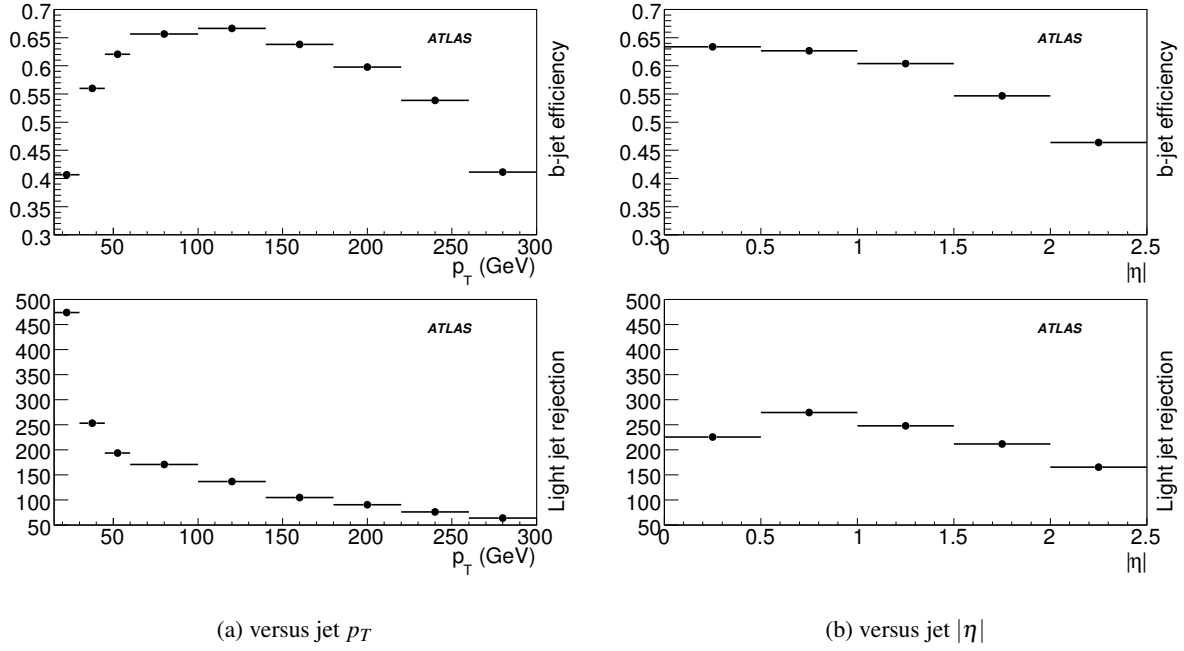


Figure 12: b -tagging efficiency and purified light jet rejection obtained with the IP3D+SV1 tagging algorithm operating at a fixed cut of 4 on the b -tagging weight, for $t\bar{t}$ events.

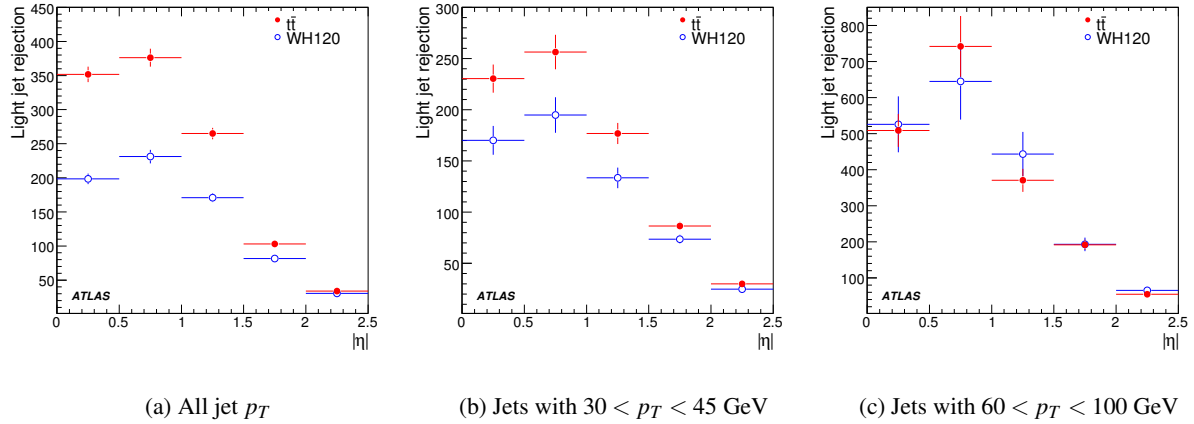


Figure 13: Rejection of light jets with purification versus jet η for the IP3D+SV1 tagging algorithm and for two different physics channels: jets from $t\bar{t}$ events and from WH ($m_H = 120$ GeV) events, for a fixed 60% tagging efficiency in each bin.

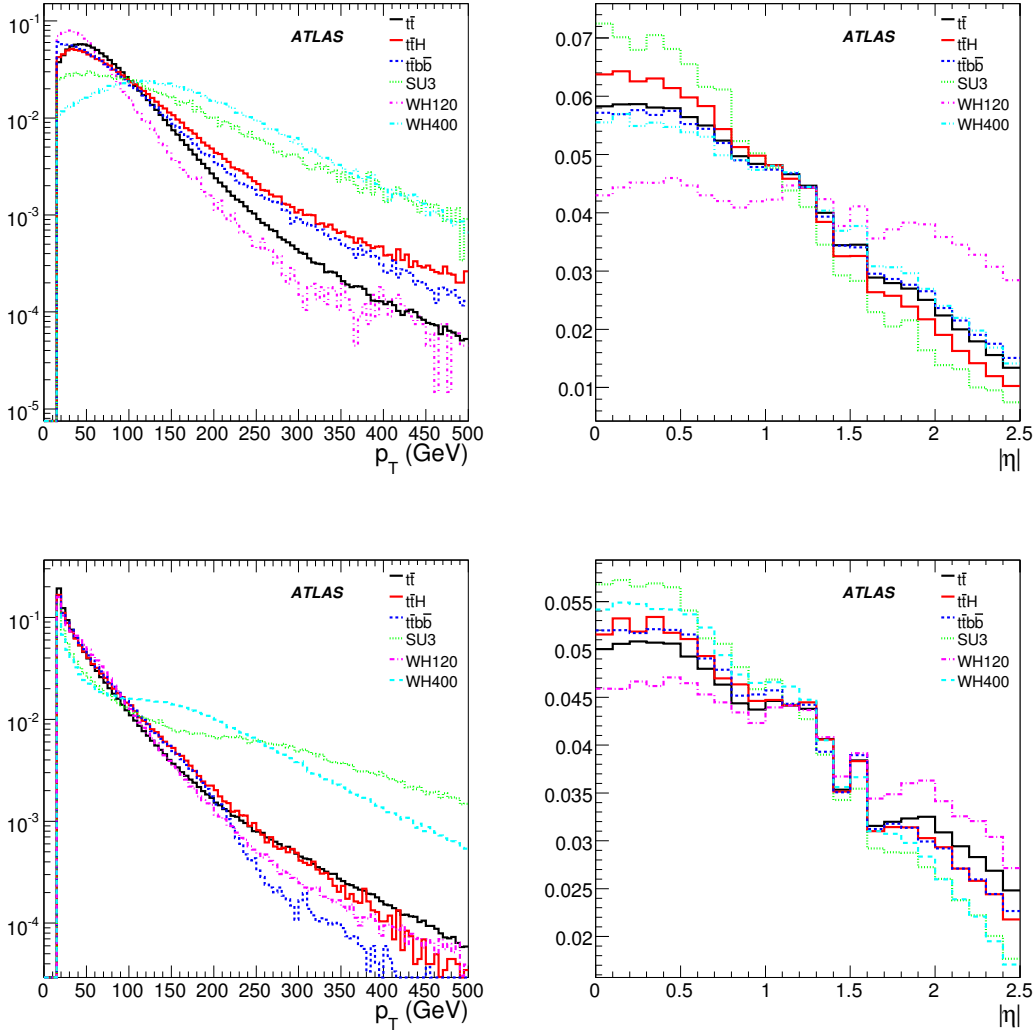


Figure 14: p_T and $|\eta|$ spectra of b (upper plots) and light (lower plots) jets for the various channels considered in this section.

a function of η when looking in bins of p_T . The remaining differences are mostly because the binning in p_T is still too large, but also because of other minor differences between the samples: for example they have been generated with different Monte Carlo generators (cf. section 6.6).

For reference, the p_T and $|\eta|$ spectra of b and light jets in the various samples used in the following are shown on Figure 14. They affect the integrated rejections for the various channels.

5.2 Simple topologies: WH channels

This first class of events illustrates the performance obtained on simple event topologies where the jet multiplicity is very low. As discussed in Section 3.3, purification is not an issue in this case.

Events from Higgs boson production in association with a W boson are interesting in this respect and are a benchmark for b -tagging performance, even though the channel itself is no longer thought to be very promising at the LHC (see however Ref. [17] for a recent re-investigation). The W decays leptonically, and there are only two jets coming from the hard process, originating from the H decay. To

study the b -tagging efficiency the decay $H \rightarrow b\bar{b}$ is simulated while here for the rejection of charmed and light jets the Higgs boson is forced to decay to $c\bar{c}$ or to the unlikely $u\bar{u}$ channel respectively.

The b -tagging performance obtained on this kind of events and for $m_H = 120$ GeV is shown in Table 1 for the tagging algorithms considered. Two typical b -tagging efficiencies were considered: 50% and 60%. For each tagging algorithm, the cuts on the weight required to achieve these efficiencies were determined over the whole sample, and then applied to estimate the rejections.

To study more energetic jets, similar physics processes have been considered for a different Higgs boson mass, $m_H=400$ GeV (again such a choice is unphysical since a 400 GeV Higgs boson would not decay to $b\bar{b}$ but is useful for these studies). The results are shown in Table 1 for the light jet rejection and in Table 2 for the c -jet rejection. The differences in performance between the two mass cases are the result of the different p_T and η spectra: the jets for $m_H = 400$ GeV are more energetic and explain most of the discrepancy, this effect being only slightly balanced by the fact that jets for $m_H = 120$ GeV are more forward. For this channel, the gain obtained with JetFitter is more visible.

5.3 Multi-jets channels: the top case

The jet multiplicity in pair-produced top quark events is much higher. In the following, the channels where at least one of the W bosons decays to leptons are considered. For the dominant lepton+jets channel, there are usually at least four jets from the hard process and extra jets from radiation. Several flavours of jets are present at the same time in the event: two b -jets from the top quarks, light jet(s) and often a c -jet from the W decaying hadronically. This increases the likelihood of having light jets contaminated with heavy flavour and also makes the labelling of jets even more ambiguous as discussed previously. The benchmark curves of jet rejection versus b -tagging efficiency are shown on Figures 15(a) and 15(b), for light jets and for several tagging algorithms. The jets of the various flavours were taken from the same sample in this case, unlike for events of the WH channels. Table 1 shows the light jet rejection achieved in $t\bar{t}$ events and in $t\bar{t}jj$ events, both samples being generated with MC@NLO+HERWIG. The latter events are $t\bar{t}$ events which were filtered in order to have at least six jets, of which four are taggable. Since the performance in those two samples is similar, they have been merged. For light jets, both the raw (without purification) and purified rejections are shown. For c -jets and τ -jets the purification does not make any significant difference. The rejection power of c -jets (Table 2 and Figure 15(c)) is naturally very limited because of the lifetime of c -hadrons and is almost independent of the physics process. Without any optimization, the b -tagging algorithms also prove to be useful for the identification of τ -jets, as shown on Figure 15(d). The small discontinuities in the curves on Figure 15, visible notably for the IP3D tagger, are due to the conjunction of a coarser binning of the underlying probability density functions for this tagger and the presence of single-track jets (notably electrons faking jets). This effect is more pronounced for the τ -jets (Figure 15(d)) where single-prong decays are abundant.

The impact on the light jet rejection of electrons faking jets can be seen in the fourth block of Table 1. Electron jets are seldom mis-tagged as b -jets, since they have usually a single prompt high- p_T track which is well-measured. In this sample, the high- p_T electrons are coming from $W \rightarrow e\nu$ or indirectly from $W \rightarrow \tau\nu$. In the fourth part of Table 1, a jet j is considered as an electron faking a jet, and therefore discarded, if it matches with a reconstructed electron candidate e : $\Delta R(e, j) < 0.1$ and $E_T(e)/E_T(j) > 0.75$.

It is interesting to notice that, despite the more complex topology of these $t\bar{t}$ events, the integrated light-jet rejection achieved is higher than for the WH ($m_H=120$ GeV) case. This is mostly because jets in $t\bar{t}$ events are more central than the ones in WH ($m_H=120$ GeV) production (cf. Figure 14), and the b -tagging performance degrades quickly at large pseudo-rapidities as seen already.

Table 1 shows the light jet rejection achieved in even more complex topologies with at least six jets. Those channels are relevant for the Higgs discovery channel $t\bar{t}H(b\bar{b})$ which requires a high b -tagging efficiency since four jets are b -tagged and the cross section is low: therefore the more typical working

Table 1: Integrated rejection of light jets (with and without purification when it applies), for various event types and for several tagging algorithms. For each case, the cut on the b -tagging weight is chosen to lead to the quoted average b -tagging efficiency ε_b over the sample considered. The quoted errors are statistical only.

	JetProb	IP2D	IP3D	IP3D+SV1	IP3D+JetFitter
<i>WH</i> ($m_H = 120$ GeV) events					
$\varepsilon_b = 50\%$	83 \pm 1	116 \pm 2	190 \pm 3	458 \pm 13	555 \pm 17
$\varepsilon_b = 60\%$	30 \pm 0	42 \pm 0	59 \pm 1	117 \pm 2	134 \pm 2
<i>WH</i> ($m_H = 400$ GeV) events					
$\varepsilon_b = 50\%$	73 \pm 1	163 \pm 3	179 \pm 3	298 \pm 7	396 \pm 11
$\varepsilon_b = 60\%$	27 \pm 0	56 \pm 1	58 \pm 1	96 \pm 1	123 \pm 2
<i>t\bar{t}</i> and <i>t\bar{t}jj</i> events					
Raw, $\varepsilon_b = 50\%$	91 \pm 0	146 \pm 1	232 \pm 2	456 \pm 4	635 \pm 7
Purified, $\varepsilon_b = 50\%$	97 \pm 0	186 \pm 1	310 \pm 3	789 \pm 10	924 \pm 13
Raw, $\varepsilon_b = 60\%$	28 \pm 0	46 \pm 0	67 \pm 0	154 \pm 1	189 \pm 1
Purified, $\varepsilon_b = 60\%$	28 \pm 0	51 \pm 0	76 \pm 0	206 \pm 1	224 \pm 2
<i>t\bar{t}</i> and <i>t\bar{t}jj</i> events, once electrons faking jets are removed					
Raw, $\varepsilon_b = 50\%$	92 \pm 0	142 \pm 1	219 \pm 1	423 \pm 4	593 \pm 6
Purified, $\varepsilon_b = 50\%$	99 \pm 0	181 \pm 1	293 \pm 2	732 \pm 10	863 \pm 12
Raw, $\varepsilon_b = 60\%$	31 \pm 0	49 \pm 0	67 \pm 0	144 \pm 1	180 \pm 1
Purified, $\varepsilon_b = 60\%$	33 \pm 0	56 \pm 0	76 \pm 0	194 \pm 1	213 \pm 2
<i>t\bar{t}H</i> events					
Raw, $\varepsilon_b = 60\%$	23 \pm 0	35 \pm 0	49 \pm 1	90 \pm 2	113 \pm 2
Purified, $\varepsilon_b = 60\%$	25 \pm 0	48 \pm 1	72 \pm 1	188 \pm 5	188 \pm 5
Raw, $\varepsilon_b = 70\%$	10 \pm 0	14 \pm 0	18 \pm 0	32 \pm 0	31 \pm 0
Purified, $\varepsilon_b = 70\%$	11 \pm 0	17 \pm 0	22 \pm 0	46 \pm 1	37 \pm 1
<i>t\bar{t}bb</i> events					
Raw, $\varepsilon_b = 60\%$	23 \pm 0	34 \pm 0	50 \pm 1	100 \pm 2	123 \pm 2
Purified, $\varepsilon_b = 60\%$	24 \pm 0	41 \pm 0	64 \pm 1	156 \pm 4	166 \pm 4
Raw, $\varepsilon_b = 70\%$	10 \pm 0	13 \pm 0	18 \pm 0	32 \pm 0	28 \pm 0
Purified, $\varepsilon_b = 70\%$	10 \pm 0	15 \pm 0	20 \pm 0	40 \pm 0	31 \pm 0
SUSY SU3 events					
Raw, $\varepsilon_b = 50\%$	66 \pm 1	140 \pm 4	162 \pm 5	246 \pm 9	328 \pm 14
Purified, $\varepsilon_b = 50\%$	68 \pm 1	161 \pm 5	183 \pm 6	290 \pm 13	375 \pm 19
Raw, $\varepsilon_b = 60\%$	24 \pm 0	50 \pm 1	55 \pm 1	89 \pm 2	110 \pm 3
Purified, $\varepsilon_b = 60\%$	25 \pm 0	53 \pm 1	58 \pm 1	99 \pm 3	117 \pm 3

Table 2: Integrated rejection of c - and τ -jets, for various event types and for several tagging algorithms. For each case, the cut on the b -tagging weight is chosen to lead to the quoted average b -tagging efficiency ε_b over the sample considered.

	JetProb	IP2D	IP3D	IP3D+SV1	IP3D+JetFitter
<i>c</i> -jet rejection for WH ($m_H = 400$ GeV) events					
$\varepsilon_b = 50\%$	7.9 ± 0.1	9.7 ± 0.1	10.7 ± 0.2	12.4 ± 0.2	12.7 ± 0.2
$\varepsilon_b = 60\%$	4.7 ± 0.0	5.7 ± 0.1	6.1 ± 0.1	6.8 ± 0.1	7.3 ± 0.1
<i>c</i> -jet rejection for $t\bar{t}$ and $t\bar{t}jj$ events					
$\varepsilon_b = 50\%$	8.4 ± 0.0	9.5 ± 0.0	10.6 ± 0.0	12.4 ± 0.1	12.3 ± 0.1
$\varepsilon_b = 60\%$	5.1 ± 0.0	5.8 ± 0.0	6.5 ± 0.0	7.4 ± 0.0	7.4 ± 0.0
τ -jet rejection for $t\bar{t}$ and $t\bar{t}jj$ events					
$\varepsilon_b = 50\%$	10.2 ± 0.1	13.9 ± 0.1	20.3 ± 0.2	45.2 ± 0.8	36.9 ± 0.6
$\varepsilon_b = 60\%$	5.1 ± 0.0	6.4 ± 0.0	8.0 ± 0.1	24.6 ± 0.3	19.3 ± 0.2

points of ε_b around 60-70% are shown. As shown in Figure 14, the p_T spectrum for the b -jets in these samples is harder than for b -jets in $t\bar{t}$ events, explaining partly the differences in performance. In addition the high b -jet multiplicity in $t\bar{t}H$ events leads to some lifetime contamination in the few light jets available in this sample. All these samples are based on PYTHIA Monte Carlo, unlike the $t\bar{t}jj$ sample which is a background for this channel as well but is based on MC@NLO+HERWIG Monte Carlo and was kept separate for this reason.

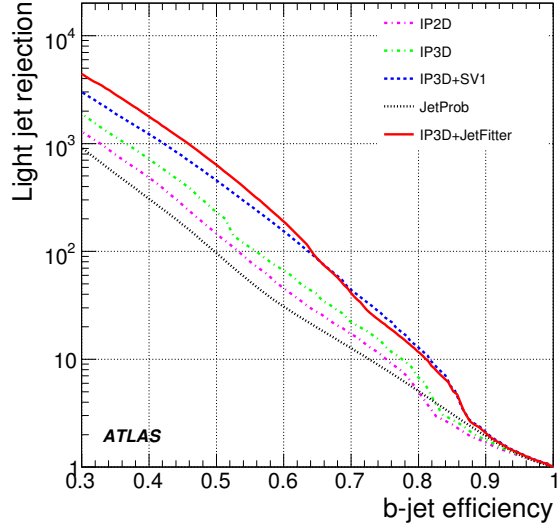
5.4 High- p_T jets: SUSY

Events from the SUSY bulk region (SU3 point, see Ref. [1]) were considered. In these events, the average taggable jet multiplicity is about 5.3 and a large number of τ -leptons are produced in the decay chain of charginos and neutralinos. There are on average 0.6 b -jets per event, with a relatively hard p_T spectrum as shown on Fig. 14: the average p_T is 144 GeV. On average about 0.6 taggable jets per event are labelled as τ , compared to 0.2 in the semi-leptonic $t\bar{t}$ channel. They are not considered as light jets. The results are shown in Tables 1 and 2: because the p_T of the jets is quite high, the light jet rejection is similar to that achieved for WH ($m_H = 400$ GeV) events but significantly worse than for the other channels.

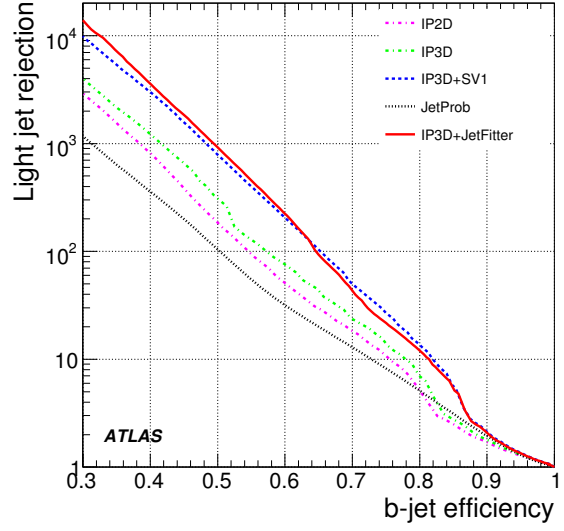
5.5 Degradation of performance at low and high p_T

At low p_T , performance is degraded mostly because multiple scattering is increased. This also holds for the high $|\eta|$ region, where the amount of material in the tracking region increases very significantly, inducing more secondary interactions. There is currently no rejection of secondary interactions found in the pixel disks, unlike in the barrel (cf. Section 4.1.3). More importantly, the increase of the extrapolation distance from the b -layer to the primary vertex at large pseudo-rapidities significantly degrades the z_0 resolution as seen in Figure 6(b).

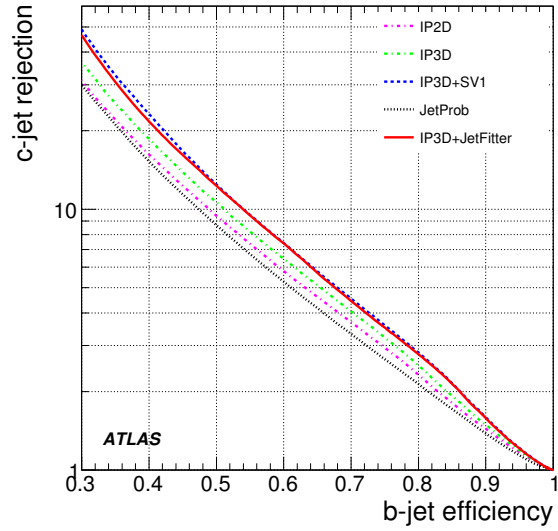
Several effects conspire to reduce the b -tagging performance as the jet p_T increases above 120 GeV. First of all, the fraction of fragmentation tracks increases with the parton transverse momentum, as shown in Figure 16, while the jet is collimated into a narrower cone: since a fixed-size cone is currently used to associate tracks to the jet this leads to a dilution of the discriminating power in b -jets. The density of tracks in the core of energetic jets challenges the pattern-recognition ability of the software and of the inner detector itself, leading to either a reduced tracking efficiency or a high level of fakes as shown in Figure 3 for jets with $E_T > 100$ GeV. Finally, for very energetic b -hadrons the Lorentz boost leads to a much enhanced decay length. The typical $c\tau$ ($\sim 450\mu\text{m}$) of b -hadrons is thus scaled by a factor



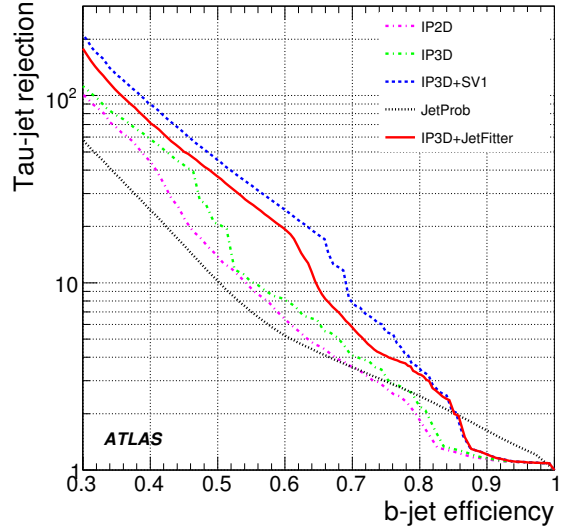
(a) Non-purified light jets



(b) Purified light jets



(c) c -jets



(d) τ -jets

Figure 15: Rejection of light jets, c - and τ -jets versus b -jet efficiency for $t\bar{t}$ and $t\bar{t}jj$ events and for all tagging algorithms: JetProb, IP2D, IP3D, IP3D+SV1, IP3D+JetFitter.

$\gamma \sim |p_B|/m_B$ which can be large. For high p_T jets, the b -hadron can decay at a rather large radius R_B , as illustrated in Table 3: close to the inner radius of the pixel detector, leading to more tracking ambiguities in the first detection layers, or even after the first pixel layer. In the latter case, the current requirement (for the IPnD and JetProb tagging algorithms) of a hit on the b -layer actually kills the signal. At very high p_T (above 500 GeV), these effects become so critical that a dedicated strategy has to be devised, as discussed in the next section. In the current simulation, the b -hadrons decaying at a radius larger than the beam-pipe or the b -layer radii do not interact with these objects, while in real events this will even more reduce the performance.

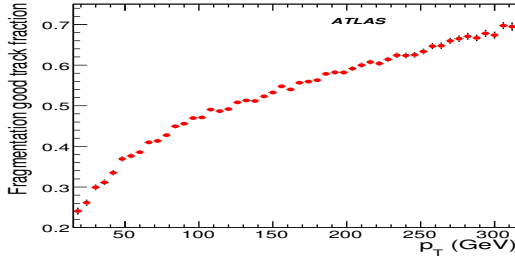


Figure 16: Fraction of selected tracks which are not from B/D decays versus jet p_T , in b -jets from WH ($m_H = 400$ GeV) events.

Table 3: Fraction of b -jets in WH ($m_H = 400$ GeV) events for which the b -hadron decays beyond the beam-pipe vacuum (first column) or beyond the b -layer (second column).

	$R_B > 2.9$ cm	$R_B > 5.1$ cm
all E_T	9.0%	2.8%
$E_T > 100$ GeV	12.2%	3.9%
$E_T > 200$ GeV	21.1%	7.9%

5.6 The case of very high- p_T jets: exotic physics

The very high p_T range is defined as jets exceeding a transverse energy of 500 GeV. Identification of such very high p_T b -jets is required for the search for heavy resonances with (predominantly) hadronic decays. A large number of exotic physics models presents signatures with very high p_T b -jets, up to a few TeV. An example is the decay $Z_H \rightarrow Zh$ in the little Higgs model [18], where $Z \rightarrow e^+e^-$ and $h \rightarrow b\bar{b}$.

In the following, three samples corresponding to the process $Z' \rightarrow q\bar{q}$, where q denotes u , b and c quarks respectively were used. The Z' mass is chosen to be 2 TeV, so that the primary partons have transverse momenta in the range from 300 GeV to 1 TeV.

The b -tagging algorithms rely particularly on the determination of the jet axis as an approximation of the b - and c -hadron flight direction. The difference between the jet pseudo-rapidity and the true b -hadron direction exhibits a narrow Gaussian core. The widths of the core range are $\sigma_\eta \approx 0.025$ and $\sigma_\phi \approx 0.010$ mrad, with a moderate dependence on jet E_T and pseudo-rapidity. Non-gaussian tails give rise to large RMS: $\text{RMS}_{\eta,\phi} \approx 0.050$. The particles from the decay of the highly boosted b -hadron are emitted at very small angles. Up to half of the tracks from the b -hadron decay lie within the azimuthal angle between the b -hadron and the jet. In these cases the sign of the impact parameter cannot be determined accurately.

The reconstruction of tracks in high p_T jets presents a series of specific challenges, as already explained in Section 5.5. Particles are associated to reconstructed jets using a $\Delta R < 0.4$ criterion. To highlight reconstruction effects only true pions reaching the outer radius of the tracker are taken into account. Particles are moreover required to originate from a well-defined vertex: either the primary vertex of the event or the b/c -hadron decay vertex. Only the first level of the efficiency, i.e. the matching to hits (Section 2.1.2), is considered. With this definition the efficiency on a reference sample of low p_T jets is close to 100% and essentially independent of the jet energy and of the distance to the jet axis. In Figure 17 the track efficiency in the Z' sample is plotted versus the jet transverse energy, for the default track reconstruction - NewTracking - and for iPatRec.

The algorithmic reconstruction efficiency for prompt tracks is only slightly degraded. Even in the very harsh environment of a 1 TeV jet, the efficiency is approximately 90%. The degradation is most

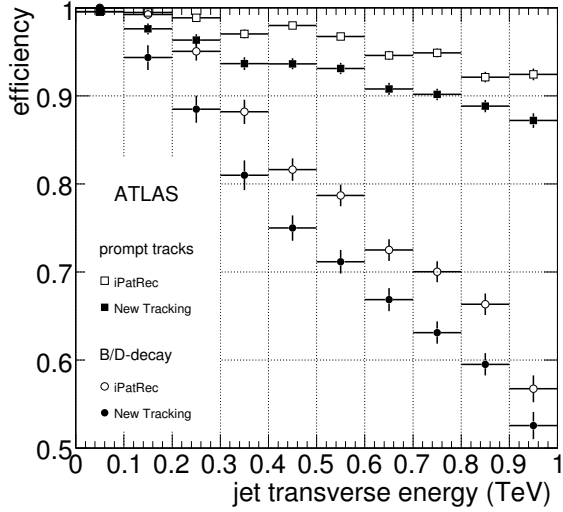


Figure 17: Algorithmic tracking efficiency for long-lived pions as a function of the jet transverse energy, for prompt tracks or tracks from b/c -hadron decays, and for two reconstruction algorithms: NewTracking and iPatRec.

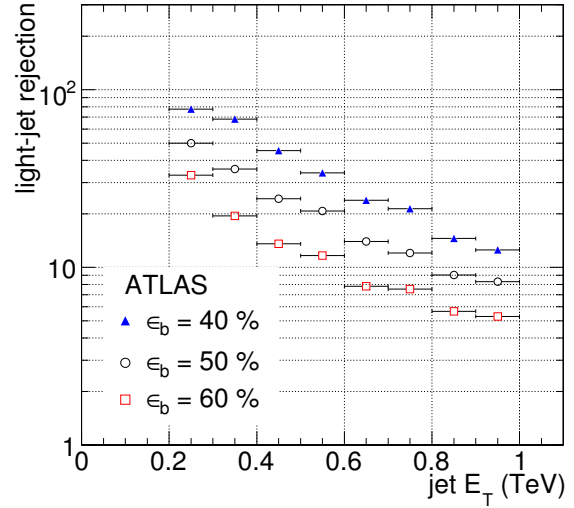


Figure 18: Raw light jet rejection for jets from $Z' \rightarrow q\bar{q}$ with $m_{Z'} = 2$ TeV, versus jet transverse energy, for the IP3D+SV1 tagging algorithm (with tracks from iPatRec). and for three b -tagging efficiencies.

pronounced in the core of the jet ($\Delta R < 0.1$). For pions originating in b/c -hadron decays a much more significant degradation of the efficiency towards high jet E_T is observed. For 1 TeV jets the efficiency is approximately 50%. The efficiency shows a strong dependence on the decay vertex radius. It is worth mentioning that NewTracking and iPatRec assign very different errors to the positions defined by large pixel or SCT clusters arising in the inner layers with such dense jets: the former assigns a very small error assuming only one particle was involved in the cluster, while the latter is assuming the opposite and assigns the maximal error (cluster width/ $\sqrt{12}$).

In very high p_T jets the probability that two or more tracks share a hit is very high ($> 10\%$), unlike what was seen for low and moderate p_T jets (cf. Figure 4). The number of shared hits per track - evaluated for individual sub-detectors, or for the complete silicon tracker - is a factor 2-3 larger in iPatRec. For particles from very displaced b/c -hadron decay vertices reconstructed with iPatRec, tracks with a shared hit in the b -layer actually outnumber the tracks with an unambiguous assignment. Again the two tracking algorithms have made opposite choices for their working point: NewTracking considers shared hits are stemming from pattern-recognition errors and tries to assign them to the best track, which is not necessarily meaningful when the cluster is really originating from several near-by particles, while iPatRec does not try to resolve the ambiguity.

The high multiplicity of fragmentation tracks, the degraded tracking efficiency, the ambiguities in hit assignment particularly in the innermost layer and the uncertainty in the impact parameter sign, all render high p_T jets a harsh environment. This may be improved though by the use of dedicated reconstruction algorithms. A priori, the current simulation should also be updated to transport properly the b -hadrons through the material. While studies have started, they are beyond the scope of this note. For the time being, a re-optimization of the default tagging algorithm parameters has been performed with the aim of improving the b -tagging performance over a large jet E_T range (from 200 GeV to 1 TeV). The cone size for the jet-to-track association was reduced to 0.2 (see Section 7.2 for a possible improvement of the current treatment), the p_T cut on tracks raised to 5 GeV and the use of tracks without a hit in the b -layer

was allowed (see also Section 7.1).

The tagging performance on iPatRec tracks is found to be on average 60% better than for the default algorithm. The choice of iPatRec to maintain high tracking efficiency inside dense jets (cf. Figure 3), even at a price of higher fake rates, seems to be instrumental in achieving better performance here. The resulting b -tagging performance is presented in Figure 18. Given the modest level of rejection achieved, a tagging efficiency of 40% is considered here. Without any tuning, the rejection level of the IP3D+SV1 tagging algorithm would be about three times worse for the same b -tagging efficiency.

In this study, the standard (low p_T) reference histograms were used for the tagging algorithms. So far, none of the methods developed to extract the calibration histograms from data has been shown to work for these very high p_T jets. Reference histograms for very high p_T jets may be extracted from Monte Carlo simulation, provided it reliably describes the data. Doing so, a modest improvement of the performance (up to 50% higher light jet rejection for the same b -tagging efficiency compared to the results shown here) can be achieved.

To conclude, b -tagging for very high p_T jets faces a series of specific difficulties. This study demonstrated that a rejection between 10 and 70 for jets with $p_T > 500$ GeV can be achieved by tuning the current algorithms. Further improvements require dedicated treatments at the clustering level (with probably a second-pass approach to break down large clusters coming from near-by particles) and at the pattern-recognition stage of the track reconstruction. On small preselected datasets, retracking with a specially optimized pattern-recognition algorithm should be possible.

6 Specific studies to characterize b -tagging performance

In this section, a few additional studies aimed at better understanding some critical aspects of the b -tagging performance are detailed. Those studies are described in a separate section because either they required specific datasets or they rely on software and/or cuts/optimizations which are different from the ones currently in use in the ATLAS software.

6.1 Impact of residual misalignments

All the studies in this note do not take into account the effect of residual misalignments. While all the samples studied were simulated with misalignments, they were reconstructed assuming a perfect knowledge of those misalignments. However, detailed studies, discussed in Ref. [7], are in progress on this subject. Two different approaches were used: residual misalignment sets and actual realignment. In the former, the events simulated with misalignments are reconstructed using the knowledge of the misalignments, but the true detector elements positions are shifted and/or rotated slightly from their actual position to mimic residual misalignments. The individual pixel modules were shifted by about $10\ \mu\text{m}$ in x and $30\ \mu\text{m}$ in y and z , and rotated by about $0.3\ \text{mrad}$. The pixel layers, disks and the whole detector were displaced by slightly smaller amounts. In this case, the light jet rejection drops by a factor 2 for the same b -tagging efficiency. For residual misalignments about half as big, the drop in light jet rejection is degraded by 40% compared to the ideal case. The reduction of residual misalignments relies on the actual alignment procedure, which is performed to obtain the new positions of the detector elements. This is the most realistic case considered so far, and includes many (but not all) systematic deformations caused by the alignment procedure itself. In this case, the light jet rejection is at most 25% lower for the same b -tagging efficiency.

6.2 Impact of the tracker material on performance

A major effort has been invested in describing accurately the material in the tracking volume of ATLAS. However, some underestimation is possible. To assess the impact of extra material on the b -tagging per-

formance, results with different geometries were compared. Extra material was added, mostly beyond the b -layer, increasing the thickness in radiation lengths by about 8% (15%) at $|\eta| \approx 0(1)$. The first noticeable effect is the degradation of the impact parameter resolution. The other effect is an increased fraction of particles undergoing interactions in the matter of the detector and producing secondary particles which can, directly or through pattern-recognition problems, fake non-prompt tracks. At a 60% b -tagging efficiency, the extra material decreases by 10% the light jet rejection power. About 60% of the loss of rejection is explained by the worsening of the impact parameter resolution, and about 40% by extra secondaries.

6.3 Impact of the pixel detector conditions

The pixel detector and notably the innermost b -layer are critical for achieving good b -tagging performance. The detector efficiency clearly affects the tracking performance but is also explicitly a key ingredient for b -tagging since the b -tagging quality cuts require that each track have at least two pixel hits of which one is in the b -layer. These pixel hit requirements are made in order to maintain the highest resolution on the impact parameter of tracks.

A single pixel inefficiency of 5% has been used to simulate the events. Measurements on the pixel staves before the detector integration gave a single pixel inefficiency below $\sim 0.3\%$ (and below $\sim 0.1\%$ for the b -layer for which the highest quality components were used). The effect of this inefficiency is especially relevant at small $|\eta|$ where half of the pixel clusters contain only one pixel. The impact of varying the fraction of randomly distributed dead pixels was studied for three tagging algorithms on a large statistics (600k events) $t\bar{t}$ sample and the results are shown in Table 4. When decreasing the fraction of dead pixels from 5% to 1%, the tracking efficiency for tracks fulfilling the b -tagging quality cuts improves by up to 2.5% absolute (around $\eta \sim 0$), leading to a relative gain in rejection of about 10%.

Table 4: Reference light jet rejection for several tagging algorithms and the relative change with various configurations of the pixel system (see text) for a 60% b -tagging efficiency in $t\bar{t}$ events. The used reference histograms were produced with the respective samples.

	IP2D	IP3D	IP3D+SV1
Reference rejection (5% of dead pixels)	54 ± 1	77 ± 2	229 ± 10
Relative change with 1% of dead pixels	+9%	+10%	+17%
Relative change with a dead half-stave on b -layer	-8%	-8%	-10%
Relative change with a dead bi-stave on b -layer	-34%	-34%	-28%
Relative change with a dead half-stave on external pixel layer	< 1%	< 1%	-1%
Relative change with a dead bi-stave on external pixel layer	-1%	-1%	-4%

More global problems, such as chip and module inefficiencies were not considered in the studies. Their effect has been studied in detail in Ref. [19] and can be very important. However, the latest measurements made right before the detector integration indicate that only one module (in the middle layer) out of 1744 is dead and that fewer than 5 chips are dead out of 27904. Besides single module failures, more dramatic failures might happen. During the pixel operation, it is thought that the two most likely sources of potential failures could be an opto-board failure, leading to half a stave (at most 7 modules) not functioning and a cooling problem implying that a whole bi-stave (26 modules) could not be used. To study those scenarios, the pixel digitization was modified to disable the corresponding modules, in either the b -layer or the external pixel layer. The impact on b -tagging performance of these two scenarios is shown in Table 4. In the case of well-identified module failures, it is clear that some recovery strategies can be used, either directly in the tracking code or at least in the b -tagging algorithm,

for instance by not requiring a hit on a dead module.

6.4 Jet algorithms

For b -tagging purposes, an accurate knowledge of the jet direction is relevant. In the first place, this direction is used to define which tracks should be associated to the jets. Then it is used to sign the impact parameter of tracks.

As mentioned earlier, all b -tagging results are given for jets reconstructed with a cone algorithm of size $\Delta R = 0.4$. Since a given physics analysis may opt for a different jet algorithm, the impact of this choice on the b -tagging performance is tested in this section. In all cases, as it is the default in the b -tagging software, only the tracks within a distance $\Delta R < 0.4$ of the jet axis were used for the tagging, even for jet algorithms which could benefit from a less geometric track-jet association such as the k_T algorithm.

Several cone sizes ΔR and size parameters R were studied for the cone algorithm and the k_T algorithm: from 0.2 to 0.8 in steps of 0.1. Finally, the b -tagging performance of the mid-point algorithm [20], an alternate jet algorithm addressing the infrared sensitivity of cone algorithms, was also checked, for two different cone size of $\Delta R = 0.4$ and 0.7. In this study, electrons faking jets were removed.

Figure 19 shows the rejection of light jets versus the b -tagging efficiency obtained with the IP3D+SV1 tagging algorithm, for several jet algorithms run on $t\bar{t}$ events. Results with and without purification are very different. With purification (Figure 19(b)) there is no significant difference in performance in the relevant range of b -tagging efficiency ($40\% < \epsilon_b < 75\%$).

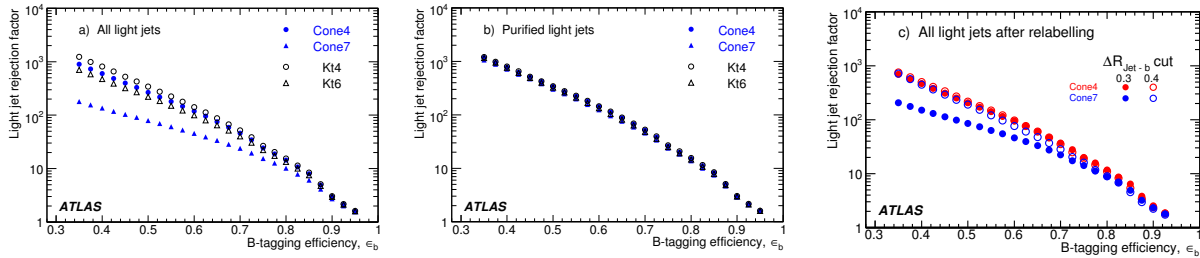


Figure 19: Rejection of light jets versus b -tagging efficiency for the IP3D+SV1 tagging algorithm applied on jets reconstructed with different algorithms: cone algorithm with size $\Delta R = 0.4, 0.7$ or k_T algorithm with parameter $R = 0.4, 0.6$. See text for the last plot.

This stability is the anticipated behavior, since only the jet direction is meaningful for b -tagging purposes and it does vary with the jet algorithm but not drastically for moderate jet p_T : the mean of the distance $\Delta R(b, jet)$ for a 50 GeV jet is 0.081 for a jet of size $\Delta R = 0.4$ and 0.095 for a jet of size $\Delta R = 0.7$ (see Ref. [9] for more details). For completeness, it should be mentioned that no differences were found between tower-based and topological cluster-based jets.

Without the purification procedure (Figure 19(a)), the results are different for different jet definitions. However, the interpretation is not straightforward. In principle broader jets could be more easily contaminated by neighbouring tracks originating from distinct partons whose showers could not be resolved: light jets for instance could be contaminated by heavy-flavour decay products. However, this effect should be marginal since the maximum track-jet distance for association is kept to $\Delta R = 0.4$ in all cases.

A better explanation is linked to the ambiguity and arbitrariness of the labelling procedure: with fewer, broader jets the assignment of partons to the jets is more ambiguous and more likely to have

$\Delta R(b, jet) > 0.3$, and thus more jets wrongly labelled as light jets. Therefore, for example, the rejection at $\varepsilon_b = 50\%$ with a cone radius $\Delta R = 0.7$ appears to be three times less than with a cone of radius 0.4, by relabelling the jets with a ΔR cut of 0.4 instead of 0.3 this difference can be virtually eliminated as seen in Figure 19(c). Thus although it appears to be more difficult to define the true flavour of a broader jet this does not necessarily exclude the choice of a cone size of 0.7 for a given analysis.

6.5 Sensitivity to the calibration of tagging algorithms

Most of the ATLAS tagging algorithms make use of an a priori knowledge to discriminate b -jets from light jets, which comes in various forms. The simplest of these ingredients is the transverse impact parameter resolution function used by the JetProb tagging algorithm to measure the compatibility of tracks with the primary vertex. The likelihood ratio tagging algorithms rely on several such distributions, with the further complication that they must be known for both the light and the b - hypothesis. In this section, the way this knowledge may affect the performance is studied. It is not currently possible to know if these settings are a good representation, both in nature and amplitude, of the differences data/Monte Carlo that will be observed, but at least they give information about the robustness of the tagging algorithms. All the results are given for the $t\bar{t}$ sample.

6.5.1 JetProb tagging algorithm

JetProb is expected to be one of the first tagging algorithms to be commissioned in ATLAS. To perform well, the resolution function (cf. Section 4.2) must be measured in data to avoid possible short-comings of the Monte Carlo simulation (resolutions and non-gaussian tails mostly). One of the major advantages of this tagging algorithm is that a priori any track from any physics process, e.g. the tracks from the first minimum-bias events, could be used to calibrate the resolution function, provided the contamination of non-prompt tracks can be kept at a very low level.

The sensitivity to this last point was studied by checking two different scenarios to select tracks in order to build the resolution functions. In all cases, only reconstructed tracks with negative impact parameter significance and fulfilling the b -tag quality cuts are used. This was done on $t\bar{t}$ events but similar or better (because of less heavy flavour contamination) results are expected for minimum-bias events. In the ideal case, the tracks are required to match to a true particle whose true origin is at the primary vertex of the event. In the realistic case, this requirement was not enforced. The distributions of the negative impact parameter significance d_0/σ_{d_0} of tracks obtained in the two cases exhibit significant differences in the tails: in the ideal case, 0.6% of the tracks have $|d_0| > 5\sigma_{d_0}$ (the RMS of the distribution is 1.3) while this fraction is 3.2% for the realistic case (RMS is 2.1). These distributions are then used as resolution functions to measure the b -tagging performance on the same events: at a 50% b -tagging efficiency, the light jet rejection with the realistic scenario is 15% lower than for the ideal case.

6.5.2 Likelihood-based tagging algorithms

For the likelihood-based tagging algorithms, the probability density functions for all the variables (cf. Section 4.1) are built for the b -jet and light jet hypotheses using Monte Carlo. The $t\bar{t}$ channel can be used in data to isolate a sample of pure b -jets from which the various distributions can be derived, using the methods described in [9]. However, more than a few hundreds of pb^{-1} of data are needed. For the light jets, it seems very difficult to isolate a pure enough sample in data. In this section, the sensitivity to the calibration is estimated using reference histograms obtained from Monte Carlo with very different settings.

First the impact of using a different tracking algorithm for the calibration (iPatRec) and for the performance measurement (NewTracking) was assessed and led to a very small variation of the rejection

power, below 10%. Another study consisted of using different detector descriptions for calibrating and testing: the two geometries compared were relatively similar, with a relative difference in the amount of material in the tracking volume of 8% (15%) at $\eta \sim 0$ (1). At most a 5% change in rejection power is seen in this case. Another issue is the sample composition of the reference histograms: using only $t\bar{t}$ events or a mix of $t\bar{t}$, WH ($m_H = 120, 400$ GeV) and SUSY events does not change significantly ($< 5\%$ relative change on rejection power) the b -tagging performance on a $t\bar{t}$ sample. However, larger effects are expected when b -tagging is run on a very different sample from the one used for the calibration.

A possible bias when building the calibration distributions and looking at performance on the same event sample was studied by dividing the sample into two. The bias on the resulting rejection factors was found to be usually negligible, and in all cases below 10%.

Finally, it was checked what statistics are needed to define the underlying histograms for the various tagging algorithms in order for results to be stable. This was checked on semi-leptonic $t\bar{t}$ events by halving a 600k event sample, building calibrations on 10k, 50k, 100k and 300k events from the first half-sample and checking the performance on the other half. To obtain a rejection level stable within 3%, 50k events are needed for all the IP and SV tagging algorithms when used in a regime where $\varepsilon_b \geq 50\%$.

6.6 Sensitivity to the Monte Carlo modelling

In the Monte Carlo modelling, several parameters can affect the ability to tag b -jets. Any effect that can change the lifetimes of the produced particles, the multiplicity of the charged tracks or the momenta of these tracks can potentially change the tagging efficiency. This modelling is not necessarily a good description of data, and in addition it is also performed differently across generators.

6.6.1 Fragmentation

First of all, various fragmentation models, describing the non-perturbative process in which quarks hadronize into colorless hadronic states, are implemented in the Monte Carlo generators. For heavy-flavour quarks, two options are available in the PYTHIA generator: the Lund-Bowler model and the Peterson fragmentation model. While the former has been found to give reasonable agreement with experimental data at LEP, SLC and HERA, the latter is currently the default in ATLAS PYTHIA productions. The impact of these various fragmentation models was investigated with six different $t\bar{t}$ samples. For the Peterson fragmentation, three samples were produced with different values for the ε_b parameter: 0.003, 0.006 (default), 0.012. For the Lund-Bowler model, the r_Q parameter was varied: 0.50, 0.75, 1.0 (default). The maximum relative discrepancy in the b -tagging efficiency was found to be around 6%, comparing the Peterson model with $\varepsilon_b = 0.012$ to the Lund-Bowler model for $r_Q = 0.5$. However, this choice of parameters is a bit extreme: the difference between the default Peterson model and the Lund-Bowler model with $r_Q = 0.75$ (which was found to fit best the OPAL and SLD data) leads to an uncertainty on the b -tagging efficiency of 1.1% for a fixed b -tagging cut leading to a b -tagging efficiency of 72%.

6.6.2 Heavy flavour production

The production fraction of various b/c -hadron species can also lead to different b -tagging efficiencies since they have different lifetimes and decay modes. The measured b -hadron fractions [15] and their values in the generators are shown in Table 5. For HERWIG, the defaults have been changed following the CDF tuning [21] by setting the CLPOW parameter to 1.2, in order to obtain a b -baryon fraction in agreement with the PDG and with PYTHIA. Another ingredient is the production of excited bottom and charm states which can give rise to soft charged pions or kaons, affecting the topology of the events: this has not been studied yet.

Table 5: Fraction (in %) of b -hadron species from the PDG (assuming $f(B_d) = f(B^\pm)$) and in PYTHIA, the default HERWIG and HERWIG tuned for ATLAS (CLPOW=1.2).

	B_d	B^\pm	B_s	Baryons
PDG	39.8 ± 1.0	39.8 ± 1.0	10.4 ± 1.4	9.9 ± 1.7
PYTHIA (ATLAS)	39.7	39.2	12.1	9.1
HERWIG	44.3	44.8	10.8	0.0
HERWIG (tuned)	39.4	39.9	10.4	10.3

The various production fractions of each type of the b -mesons were varied according to the measured errors from PDG 2006 [15], and the impact on a simulated PYTHIA $t\bar{t}$ sample was studied by a re-weighting technique. This source of systematics can be safely neglected since the net effect is an uncertainty below the per mil level on the tagging efficiency.

6.6.3 b -hadron lifetimes and decays

The uncertainty on the lifetime of the various b hadrons was also studied and found to give rise to an uncertainty of 0.3% for a b -tagging efficiency of 72%.

The uncertainty in the charged track multiplicity of b hadron decays was estimated by comparing PYTHIA with measurements from LEP [22]. The resulting uncertainty on the b -tagging efficiency was found to be 0.9%.

6.6.4 Heavy flavour decays with EvtGen

The two event generators used in ATLAS to fragment and decay particles, PYTHIA and HERWIG, implement different algorithms to simulate the decays of generated particles, using their own decay tables to specify decay modes and branching fractions. The sophistication of the decay simulation and the scope of decay tables vary considerably between generators. For B meson decays, arguably the most detailed simulation is currently provided by EvtGen [23].

Since the b -tagging performance on Monte-Carlo samples may depend on details of simulated particle decays, such as the charged particle multiplicity or the spatial distribution of secondary decay vertices in B decays, the impact of using EvtGen as a decayer instead of PYTHIA was studied on $t\bar{t}$ events. For this study, a decay file for inclusive decays was assembled based on the latest (as of summer 2005) version of the decay files used by the BaBar and CDF experiments. For decay channels where experimental data is available, branching fractions were taken from PDG [15], while the remaining decays are simulated generically with JETSET.

For this study, two specific samples were generated: the first $t\bar{t}$ sample was generated by PYTHIA and the decays were handled by PYTHIA. The second one was generated in the same way except that particle decays were simulated by EvtGen. As expected, changing the particle decay simulation leads to small differences in some distributions of generator-level quantities. For example, the mean multiplicity of charged pions in decays of B^\pm and B^0 mesons, not including decays of long-lived weakly decaying strange particles such as K_s^0 and Λ , is 3.89 (RMS 2.19) with PYTHIA and 3.62 (RMS 2.13) with EvtGen. The average multiplicity obtained by EvtGen agrees well with the experimental value of 3.58 ± 0.07 [15].

The b -tagging weight distributions obtained with the two generators are thus slightly different. For a fixed b -tagging weight cut, the b -jet efficiency varies by about 1%. Tuning the cut to keep the same b -tagging efficiency in both samples leads to non-negligible changes in the light jet rejection: it decreases by about 5% to 15% when EvtGen is used, depending on the tagging algorithm and chosen b -jet efficiency.

7 Specific studies for improving b -tagging performance

In this section, three studies aiming at improving the b -tagging performance are presented. Most of them rely on specific software developments.

In the first part, new track categories are defined to make a better use of the slight differences in e.g. impact parameter resolution that such tracks may exhibit. A second study evaluates the potential gain by varying the way tracks are associated to jets. The last study shows the improvement obtained when combining several tagging algorithms in a multivariate approach. Those studies were done independently and no attempt was made yet to combine them. It is worth noting that some approaches advocated in the first two studies are expected to be highly correlated.

7.1 Improving performance with track categories

Grouping the tracks used for b -tagging in several categories has been discussed in Section 4.1.5. Dedicated treatment for *Shared* tracks (cf. 2.1.3) is already implemented and used in the current b -tagging software. Using dedicated probability density functions for the *Shared* tracks improves the light jet rejection by 23% (7%) for a b -jet tagging efficiency of 50% (60% respectively) in $t\bar{t}$ events. This is the default treatment in the software. This effect is sample-dependent and is more important for samples with high jet multiplicities and energetic jets, which tend to be more collimated.

Using additional track categories to improve the b -tagging performance is being further investigated. A possible interest of the track categories is to try to loosen the track quality cuts and therefore gain in efficiency without diluting the discrimination power of the good tracks. As discussed in Section 2.1.2, requiring each track to have a hit on the b -layer leads to an absolute loss in efficiency of about 2.5%, which is not negligible for b -tagging purposes where few tracks are available. Thus one attempt consisted of trying to keep tracks with no b -layer hit in a special category. About 4% of the tracks in jets from $t\bar{t}$ events which pass the rest of the b -tagging selection fall in this category.

The categories could also be used to deal with the non-Gaussian resolution tails and the imperfect treatment of the matter in the tracking error estimation process. A priori some fraction of these effects would be better treated from first principles directly in the tracking, but the experience with previous experiments shows that this is difficult in practice and therefore ad-hoc treatments may be justified. The natural variables to partition tracks are p_T (actually p for multiple-scattering) and pseudo-rapidity (since the material is very non-uniform in η).

Finally, another potential use of track categories was studied: in b -jets, the fragmentation tracks accompanying the b -hadron decay products are prompt and should therefore have different distributions of the discriminating variables. Tracks with $p_T(\text{track})/p_T(\text{jet}) < 0.04$ were defined as fragmentation tracks since those are in principle softer and were put in a special category. In typical b -jets from the $t\bar{t}$ sample, this cut selects 13% of the tracks. Some correlation with the treatment in p_T bins (previous case above) is expected.

The use of these categories brings some improvement to the light jet rejection, as high as 60% for the binning in track p_T which is the most powerful way of partitioning tracks. The improvement for some categories depends on the sample: the dedicated category for fragmentation tracks for example is more helpful for the WH ($m_H = 400$ GeV) sample where the actual decay products of the b -hadron are very collimated and the fixed-size cone for associating tracks to the jets brings in a larger fraction of prompt tracks.

Various ways to combine all the new categories were investigated. The gain in rejection brought by the best combinations, using 11 partitions formed with the aforementioned categories, ranges from 20% to 70% depending on the sample and the tagging algorithm, for a 60% b -tagging efficiency.

7.2 Optimizing the track-to-jet association

As the jet transverse momentum increases, its particles are collimated into a narrower cone. But currently all tracks within $\Delta R < 0.4$ of the jet axis are associated with the jet, regardless of its momentum. For a 300 GeV b -jet, only 30% of the tracks associated to the jet comes from the b -hadron decay products, as shown in Figure 16. Therefore at high p_T the b -tagging discriminating power is diluted since a larger fraction of the tracks in the jet may be picked up from environmental contamination: underlying event, pile-up or neighbouring jets in busy events. An alternative track-to-jet association has been studied, with a ΔR cut varying with the jet p_T : $\Delta R < f(p_T)$. Based on the distribution of $\Delta R(\text{jet}, \text{track})$ for tracks originating from b -hadron decays in b -jets, a functional form $f(p_T)$ has been chosen which ensures that 95% of the tracks from b -hadron decays in these events are associated to the jet for any jet p_T in the range [15, 500] GeV.

The impact on b -tagging performance of using this association instead of the standard one is checked on $t\bar{t}$ events. The relative improvement on the overall raw light jet rejection is 46% (7%) for respectively a b -tagging efficiency of 50% (60%). This improved treatment actually affects only the non-pure jets (22% of the light jets in this sample), for which the rejection triples for $\epsilon_b = 50\%$ and doubles for $\epsilon_b = 60\%$. Obviously the fraction of non-isolated jets and therefore the possible gain with this method are sample-dependent.

7.3 Combining tagging algorithms with boosted decision trees

Several multi-variant techniques exist that can combine different b -tagging algorithms into a single classifier for discriminating b -jets from light jets. We investigated boosted decision trees (BDT).

BDT can be applied to any classification problem and their use for combining several b -tagging algorithms into a single classifier for discriminating b -jets from light jets was investigated [24]. In this study, a BDT classifier was optimized on a training sample containing b -jet and light jet patterns extracted from WH ($m_H = 120$ GeV, $H \rightarrow b\bar{b}$ or $H \rightarrow u\bar{u}$) and $t\bar{t}$ samples. The following input variables were used: the weight from the IP3D tagging algorithm, the three variables on which the SV tagging algorithms are based (cf. Figure 9), the number of tracks associated with the secondary vertex, the weights of the soft muon and soft electron tagging algorithms, the largest transverse and longitudinal impact parameter significances and transverse momentum of the tracks in the jet, the jet transverse momentum and the number of tracks in the jet. For a fair comparison with IP3D+SV1, a BDT classifier with only the first four variables was also studied. The predictive power of the classifier was estimated using a distinct sample of b -jet and u -jet patterns (test sample).

Table 6: Rejection of light jets given by IP3D+SV1 and the boosted decision tree for WH (with $m_H=120$ GeV) and $t\bar{t}$ events, for fixed b -tagging efficiencies of 50% and 60% in each sample.

	IP3D+SV1		BDT 4 variables		BDT 12 variables	
	$\epsilon_b = 50\%$	$\epsilon_b = 60\%$	$\epsilon_b = 50\%$	$\epsilon_b = 60\%$	$\epsilon_b = 50\%$	$\epsilon_b = 60\%$
WH ($m_H=120$ GeV)	529 ± 40	155 ± 6	682 ± 79	189 ± 12	762 ± 93	201 ± 13
$t\bar{t}$	393 ± 16	143 ± 3	484 ± 34	161 ± 6	563 ± 42	187 ± 8
$t\bar{t}$ after purification	720 ± 42	205 ± 6	808 ± 73	226 ± 11	1021 ± 103	278 ± 15

Table 6 compares the rejection of light jets given by the two BDT configurations and the likelihood ratio weight IP3D+SV1 for WH events and $t\bar{t}$ events. It shows that with the same variables the BDT outperforms IP3D+SV1 by 10 to 30%. When using additional information, including the soft lepton tagging, the light jet rejection on both event topologies increases by about 50% with respect to IP3D+SV1. For these results, the training and test samples were based on similar events (WH or $t\bar{t}$). When train-

ing the BDT on WH events and using $t\bar{t}$ events for the test sample, the gain in rejection compared to IP3D+SV1 is lower but still interesting ($> 20\%$).

8 Measuring b -tagging performance in data

For analyses using b -tagging, the estimation of the backgrounds from well-known Standard Model processes requires knowledge of the tagging and mis-tagging efficiency for the various flavours of jets with high accuracy. The quality of Monte Carlo simulation of these properties is unknown and therefore strategies must be developed to measure the tagging and mis-tagging efficiencies directly in data.

8.1 b -tagging efficiency

Several strategies for measuring the b -tagging efficiency directly in data are investigated in detail. The relative precision they permit has been estimated for a typical b -tagging efficiency of 60%.

The first approach, described in Ref. [8], relies on a sample of the abundantly produced dijet events, in which one of the jets contains a muon. A muon+jet trigger has been conceived and proposed for this purpose. Two methods, also employed at the Tevatron, are used to estimate the b -content of the dijet sample. The p_{Trel} method is based on templates of the muon p_T relative to the jet axis. The templates are derived from Monte Carlo events for the three types of jets: bottom, charm and light (the latter will eventually be derived from data). The so-called System 8 method uses two samples of differing bottom quark content and two uncorrelated tagging algorithms, typically the soft muon one and the tagging algorithm to be calibrated, to form a system of equations from which the b -tagging efficiency can be extracted. Using 50 pb^{-1} of data, a detailed p_T - or η -dependent calibration curve could be derived with the p_{Trel} method and with System 8. Since it is expected that the systematic uncertainties will dominate rapidly the total error for these methods, a careful study of systematics errors has to be done which was not fully completed for this note: the systematic errors studied so far indicate that it should be possible to control the absolute error on ϵ_b to 6%. Currently the two methods are proven to work well for jets below a p_T of 80 GeV.

The second approach, discussed in Ref. [9], makes use of $t\bar{t}$ events and is complementary: a little more data is needed but the tagging efficiency of higher p_T jets can be measured. Two distinct ways are described: by counting the number of selected $t\bar{t}$ events with one, two or more b -tagged jets, or by studying directly the output distributions of b -tagging algorithms on samples of b -jets pre-selected by several methods (topological, kinematic or likelihood selection). The counting method allows measurement of the integrated b -tagging efficiency with a relative precision of $\pm 2.2(\text{stat.}) \pm 3.5(\text{syst.})\%$ in the lepton+jets channel and $\pm 3.7(\text{stat.}) \pm 2.7(\text{syst.})\%$ in the di-lepton channel for 100 pb^{-1} of data. For the topological selection, 200 pb^{-1} of data are needed and allow a relative precision of $\pm 7.7(\text{stat.}) \pm 3.2(\text{syst.})\%$.

8.2 Mis-tagging rates

Measuring the mis-tagging rate in data is more difficult and under study. The main approach is to use the negative tags (either in impact parameter or decay length) which describe the effects of a limited resolution on a priori prompt tracks and then to correct for long-lived particles (K_s , Λ , etc), material interactions and heavy flavour jets which are negatively tagged. So far no study in ATLAS has estimated the accuracy with which fake rates can be measured. However, based on the Tevatron experience, it seems that a 10% relative error could be achievable with 100 pb^{-1} .

8.3 Extracting reference distributions from data

For the likelihood tagging algorithms, reference distributions for light and b -jets are needed. In the case of b -jets, they could in principle [25] be measured in data from a pure sample of b -jets using the techniques developed for the b -tagging efficiency estimation in $t\bar{t}$ events. As shown in Ref. [9], the various distributions can be checked in data with a few hundred pb^{-1} . However, much more data is needed to extract multi-dimensional likelihoods. For light jets, it seems difficult to extract from data a sample with sufficient purity. In any case, a Monte Carlo accurately describing the data is also needed to extrapolate those reference distributions to ranges where they certainly can not be measured, the very high- p_T regime for instance.

9 Conclusions and outlook: realistic performance in first data

The first tagging algorithm to be commissioned with real data is expected to be (besides the track counting method) JetProb, using the tracks from any kind of events to define its resolution function. For a b -tagging efficiency of 60%, a light jet rejection of around 30 could be achieved with this tagging algorithm but further improvements are expected. The soft lepton tagging algorithms will be commissioned at the same time, leading to higher rejection levels when considering semi-leptonic b -jets. Once the quality of the Monte Carlo simulation is checked and better understood with data, a tagging algorithm relying on Monte Carlo templates for b and light jet hypotheses such as IP3D can be used, perhaps doubling the rejection power. The commissioning of the tagging algorithms relying on secondary vertexing may take more time, but tagging algorithms like SV1 should quadruple the initial rejection level, bringing it above 100. Finally the combination of the ultimate JetFitter tagging algorithm and the various improvements described in this note should permit a rejection of 200, or more interestingly to maintain a rejection of 100 at a higher b -tagging efficiency, around 70%.

Those estimates do not take into account the impact of residual misalignments in the tracker. However, a first study has been performed in which the actual alignment procedure was run. This is the most realistic study so far, and includes many systematic deformations caused by the alignment procedure itself. It concludes that the mis-tagging rate is at most 30% lower for the same b -tagging efficiency with a realistic early detector alignment.

The impact of several other effects on the b -tagging performance has been studied in this note. In the simulation used, the fraction of dead pixels is overestimated. Using 1% instead of 5% dead pixels improves the light jet rejection by about 10%. Relative improvement in the light jet rejection, from 10% in typical $t\bar{t}$ events to 50% for high- p_T samples, could be achieved with a tuning of the tracking in jets. The sensitivity to the accuracy of the passive material description in the simulation could be quite dramatic. However, a large effort has been made in describing accurately the material in the tracking volume. If an 8% to 15% discrepancy between Monte Carlo and data would remain, the impact on the b -tagging performance is a 10% relative change in light jet rejection power.

New ideas to improve the performance have been studied: the generalization of the track categories could bring a 10% to 60% improvement, optimizing the track-to-jet association could help significantly in busy events and finally the use of multivariate techniques such as BDT could lead to gains in the range 10-50%. However, all these potential gains only apply to some tagging algorithms, some p_T region, etc. Furthermore some correlations are expected among them. Therefore it will be interesting to assess the net impact of these improvements when they are all available, used simultaneously and optimized.

Finally, detailed studies have shown that the b -tagging efficiency can be measured directly in data using dijet or $t\bar{t}$ events. With 100 pb^{-1} , a relative precision of about 5% can be achieved for b -jet efficiency. The accuracy with which the mis-tagging rates can be measured deserves more study, however, a 10% precision seems feasible based on the Tevatron experience.

References

- [1] ATLAS Collaboration, *Supersymmetry Searches*, this volume.
- [2] ATLAS Collaboration, *Top Quark Mass Measurements*, this volume.
- [3] ATLAS Collaboration, *Search for $t\bar{t}H(H \rightarrow b\bar{b})$* , this volume.
- [4] ATLAS Collaboration, *Vertex Reconstruction for b -Tagging*, this volume.
- [5] ATLAS Collaboration, *Soft Muon b -Tagging*, this volume.
- [6] ATLAS Collaboration, *Soft Electron b -tagging*, this volume.
- [7] ATLAS Collaboration, *Effects of Misalignment on b -Tagging*, this volume.
- [8] ATLAS Collaboration, *b -Tagging Calibration with Jet Events*, this volume.
- [9] ATLAS Collaboration, *b -Tagging Calibration with $t\bar{t}$ Events*, this volume.
- [10] ATLAS Collaboration, *HLT b -Tagging Performance and Strategies*, this volume.
- [11] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 (2008) S08003.
- [12] ATLAS Collaboration, *The Expected Performance of the ATLAS Inner Detector*, this volume.
- [13] ATLAS Collaboration, *Jet Reconstruction Performance*, this volume.
- [14] ALEPH Collaboration, *A precise measurement of $\Gamma_{Z \rightarrow b\bar{b}}/\Gamma_{Z \rightarrow \text{hadrons}}$* , Phys. Lett. **B313**, (1993) 535; D. Brown, M. Frank, *Tagging b -hadrons using track impact parameters*, ALEPH-92-135.
- [15] S. Eidelman *et al.*, Phys. Lett. **B 592**, 1 (2004), and 2005 partial web update for the 2006 edition.
- [16] T. Bold *et al.*, *Pile-up studies for soft electron identification and b -tagging with DC1 data*, ATL-PHYS-PUB-2006-001.
- [17] J. M. Butterworth, A. R. Davison, M. Rubin, G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, Phys. Rev. Lett. 100, 242001 (2008).
- [18] G. Azuelos *et al.*, *Exploring Little Higgs Models with ATLAS at the LHC*, hep-ph/0402037; SN-ATLAS-2004-038, S. Gonzalez de la Hoz, L. March and E. Ros, *Search for hadronic decays of Z_H and W_H in the Little Higgs model*, ATL-PHYS-PUB-2006-003.
- [19] S. Corréard *et al.*, *b -tagging with DC1 data*, ATL-PHYS-2004-006.
- [20] A. Cheplakov, A. S. Thompson, *MidPoint algorithm for jet reconstruction in ATLAS*, ATL-PHYS-PUB-2007-007.
- [21] J. Lys, private communication.
- [22] The LEP/SLD Heavy Flavour Working Group, *Final input parameters for the LEP/SLD heavy flavour analyses*, LEPHF-2001-01.
- [23] D. Lange and A. Ryd, <http://www.slac.stanford.edu/~lange/EvtGen/>; D. Lange, Nucl. Instr. Meth. **A 462** (2001) 152.

- [24] J. Bastos, *Performance of boosted decision trees for combining ATLAS b -tagging methods*, ATL-PHYS-PUB-2007-019.
- [25] S. Corréard, *b -tagging calibration and search for the Higgs boson in the $t\bar{t}H$ channel with ATLAS*, PhD thesis (in french), 2005, Université de la Méditerranée.