



ATLAS Note

ATL-PHYS-PUB-2017-011

16th June 2017



Secondary vertex finding for jet flavour identification with the ATLAS detector

The ATLAS Collaboration

A high performance identification of jets, produced via fragmentation of bottom quarks, is crucial for the ATLAS physics program. These jets can be identified by exploiting the presence of cascade decay vertices from bottom hadrons. A general vertex-finding algorithm is introduced and its application to the search for secondary vertices inside jets is described. Kinematic properties of the reconstructed vertices are used to construct several b -jet identification algorithms. The features and performance of the secondary vertex finding algorithm in a jet, as well as the performance of the jet tagging algorithms, are studied using simulated $pp \rightarrow t\bar{t}$ events at a centre-of-mass energy of 13 TeV.

ATL-PHYS-PUB-2017-011
18 June 2017



1 Introduction

The ability to identify jets containing b -hadrons (b -jets) is important for the physics program of the ATLAS experiment [1] at the LHC. During the fragmentation process a b -quark forms a b -hadron which decays via the electroweak interaction, and thus a b -jet contains charged tracks coming from the decay of the b -hadron and tracks produced in the b -parton showering. The dominant decay modes of a b -hadron are characterized by a transition of a b -quark to a c -quark, followed by the c -hadron decay. b -hadrons have relatively long lifetimes (e.g. 1.5 ps for the B^0 meson) and can travel a few millimetres before they decay. Charmed hadrons lifetimes are smaller, e.g. 1.0 ps for the D^\pm meson and 0.4 ps for the D^0 meson, but they can also travel significant distances at high energies. The displaced vertices resulting from the b -hadron and subsequent c -hadron decays provide a strong signature for a b -quark jet. These vertices can be reconstructed explicitly and their properties can be used to construct efficient b -jet identification algorithms (b -tagging), that are able to separate jets likely to contain b -hadrons from jets containing light or c -hadrons.

This note describes a high-performance algorithm to find secondary vertices in a jet, and a method by which the presence of a reconstructed vertex in a jet and kinematic features of the tracks associated to the vertex can be used for the b -jet identification. The Secondary Vertex Finder (SVF) algorithm reconstructs the displaced b - and c -hadron decay vertices, the secondary vertices (SV), inside a jet. The list of reconstructed secondary vertices and their properties are propagated to the SV-based b -tagging algorithms. The vertexing algorithm [2] and the resulting capability to tag b -jets [3] were successfully exploited by the ATLAS Collaboration for data analysis in Run I. Before the start of Run II, the SVF algorithm was improved to reflect the upgraded ATLAS detector. In particular a new innermost layer in the tracking system was installed, the Insertable B-Layer [4], and the vertex reconstruction efficiency for high transverse momentum (p_T) and high- $|\eta|$ ¹ jets was increased. The fake-vertex rate was significantly reduced, as well. The performance of the Run II version of the algorithm has been briefly presented in Refs. [5, 6]. This note provides details about the SVF design, and shows the current performance of the SVF-based b -tagging algorithms.

The note is organized as follows: Section 2 describes the general algorithm for finding secondary vertices. Section 3 describes the implementation of the algorithm in ATLAS, the selection criteria and the reconstruction of secondary vertices. Section 4 describes the application of the SVF algorithm as a part of SV-based b -tagging algorithms. The exploited vertex information and b -taggers' design are presented. Section 5 describes the performance of the vertex finder algorithm, the properties of the reconstructed secondary vertices, and the performance of the secondary-vertex-based b -tagging algorithms, including an illustration of the selection criteria, the properties of secondary vertices, the reconstruction efficiencies of the secondary vertex finders, and finally the performance of the secondary-vertex-based b -tagging algorithms. The performance is studied in simulated $pp \rightarrow t\bar{t}$ events at a centre-of-mass energy of 13 TeV.

¹ ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z -axis along the beam pipe. The x -axis points from the IP to the centre of the LHC ring, and the y -axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z -axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln \tan(\theta/2)$.

2 General vertex finding algorithm

For the reconstruction of a vertex consisting of n tracks, it is assumed that the n tracks pass very close to a common space point (the vertex position). Any two tracks out of the n tracks should be close enough to each other in order to form a two-track vertex near the common n -track vertex position. Thus, any n -track vertex can be represented as a union of two-track vertices, without a loss of generality.

The vertex finding problem can be formulated as follows: find all (interesting) vertices with an unspecified number of participating tracks from a given set of N tracks. The representation of n -track vertices as the unions of the two-track vertices allows this problem to be solved with a simple two-step algorithm:

1. Find all two-track vertices.
2. Convert the set of the two-track vertices into a multi-track vertex set by merging the two-track vertices which are close in space.

This algorithm is fast — it requires only $O(N^2)$ vertex position estimations (mainly simple two-track vertices). This can be compared with e.g. $O(N^4)$ vertex position estimations which are needed when trying to find a 4-track vertex directly.

A more important advantage of this algorithm for the application to particle physics problems comes from the observation that the origin of many of the two-track vertices can be identified. In a typical particle physics experiment environment there are several sources of vertices, such as hadronic interactions in detector material, photon conversions, V^0 decays, etc. Sometimes in a dense track environment (e.g. around a jet axis or close to the primary interaction vertex of the event) some prompt tracks produced in the primary vertex occasionally become close to each other in space far from the true production vertex, so that the vertex reconstruction algorithm considers them as coming from a common displaced vertex (i.e. a fake secondary vertex). Usually only one vertex type is of interest, e.g. b -hadron decays, and other types are considered as nuisance vertices, especially the fake ones. The present algorithm allows for the indication and eventual rejection of nuisance vertices immediately after the two-track vertex set is reconstructed. Thus the complex multi-track vertex reconstruction stage, and the algorithm as a whole, become more robust and clean. Finally, the vertex finding algorithm for particle physics problems is the following:

1. Find all two-track vertices.
2. Identify and reject as many two-track vertices as possible that are unrelated to the vertices of interest.
3. Convert the set of cleaned two-track vertices into a multi-track vertex set by merging the two-track vertices which are close in space.

This algorithm works well in a sparse track environment with well separated vertices. Problems are encountered when using such an algorithm in dense environments, as the possibility of a single track being assigned to multiple vertices in the two-track vertex set increases. If the given track set contains a single n -track vertex or all such vertices are sufficiently far from each other, the merging of the close two-track vertices usually produces a resolved vertex set, where every track is assigned to a single n -track vertex. However, in the presence of fake two-track vertices, that are inevitable in dense environments, the simple vertex-merging procedure cannot guarantee a unique track-to-vertex assignment anymore. To

resolve this issue, each multiply-assigned track in the final n -track vertex collection should be detached from all vertices except for the one with the smallest track-vertex association χ^2 .

This approach would be an effective method for solving many problems requiring (multi-)vertex reconstruction, for example pileup vertex finding, b -tagging, converted photon finding, detector material studies using hadronic interactions, etc. The number of tracks which can be input is largely limited by the CPU consumption of the algorithm.

3 Secondary vertex finder implementation in ATLAS

3.1 Description

The ATLAS implementation of the SVF is primarily designed to be used for jet flavour identification. Therefore only tracks inside a given jet cone are considered when searching for secondary vertices. The p_T -dependent jet cone size is defined by a ΔR distance from the jet axis, where $\Delta R = \sqrt{(\eta_{\text{jet}} - \eta_{\text{track}})^2 + (\phi_{\text{jet}} - \phi_{\text{track}})^2}$, and η is the pseudorapidity and ϕ is the azimuthal angle. In order to reduce the influence of the increasing number of tracks in high- p_T jets on the b -tagging efficiency, ATLAS uses a p_T -dependent cone for the track selection in jets [3]. The cone size changes from 0.45 at jet $p_T = 20$ GeV to ~ 0.24 at jet $p_T \geq 2$ TeV. The inputs of the algorithm are the jet axis direction, the primary vertex position of the event and a list of tracks associated to the jet. If successful, the algorithm returns one or more reconstructed secondary vertices, the list of tracks attached to each of them and a set of secondary vertex kinematic features which can be exploited by the subsequent jet tagging algorithms.

The SVF works as a Single Secondary Vertex Finder (SSVF), trying to reconstruct a single secondary vertex per jet. For a b -jet containing both b - and c -hadron decay vertices, the SSVF merges these vertices into a common single vertex if they are close, or reconstructs the vertex with the largest track multiplicity if they are far apart. The finite resolution of the ATLAS tracking detector does not allow the cascade decays of b - and c -hadrons to be resolved in every jet. Quite often the b - or c -vertex contain just one reconstructed track and cannot be reconstructed alone. Therefore the search for a single secondary vertex in a b -jet is justified and the vertex that is found should preserve most of the information necessary for the b -tagging. In addition the SSVF approach has a small probability of finding a (fake) vertex when there are no real secondary vertices in the jet. This feature is crucial for the rejection of light jets in the b -tagging.

The SSVF algorithm selects tracks fulfilling strict quality requirements for the secondary vertex search. Although this is not mandatory, the algorithm imposes a cut on the significance of the 3D impact parameter of a track to the event primary vertex (PV). Additional requirements on the sum of the two track impact parameter significances and on the SV-PV distance significance are imposed in the two-track vertices search. This is done to reduce combinatorics in the secondary vertex search and to remove fake secondary vertex candidates which are very close to the PV. These requirements are tuned to provide the highest b/c -hadron vertex reconstruction efficiency with the smallest fake rate.

3.2 Cleaning of two-track vertices

The straightforward secondary vertex reconstruction based on the closeness of two tracks in the space results in a large fraction of fake vertices in an environment with high track density (e.g. in the core of a

jet or in events with very high pileup). Such fake vertices can be suppressed by exploiting the following observation: any secondary track should not have associated detector hits with radii smaller than the radius of the SV found by the algorithm. This observation is used to remove fake vertices by requiring that the hit patterns of the tracks associated to a two-track vertex are consistent with the position of the reconstructed secondary vertex inside the tracking detector.

Significant numbers of secondary vertices are produced by interactions of primary hadrons with detector material, photon conversions and decays of long-lived particles (K_s and Λ). These are real vertices, but they constitute a significant background to the decays of b - and c -hadrons, therefore they need to be eliminated from the b -jet identification process. To achieve the highest purity of the reconstructed b - and c -hadron decay vertices several different methods have been developed.

V^0 decay vertices are identified by their reconstructed invariant masses. The masses of particles (proton, pion) are assigned to the same set of tracks and the vertex invariant mass is then computed. The invariant mass spectra of $\pi^+\pi^-$ and $p\pi$ are used to identify the peaks due to K_s and Λ^0 decays. In case of the $p\pi$ hypothesis the proton mass is assigned to the track with the highest p_T , while the pion mass is assigned to the lowest p_T track. The vertices in the mass peak windows are refitted using the corresponding mass constraints. The impact parameters of the combined V^0 tracks are calculated with respect to the event primary vertex. If the obtained impact parameter of the V^0 track is small, the two charged tracks coming from this vertex are marked as bad tracks and do not participate in the subsequent combined vertex search. Conversion vertices are identified as a peak with low invariant mass when assuming the e^+e^- track hypothesis.

Hadronic interactions with the detector material are rejected based on comparison of the vertex radius with radii of cylindrical material layers. The nuisance two-track vertices produced by hadronic interactions are removed from further consideration if their radii are consistent within the detector material position, as shown in Figure 2. As true b - and c -hadron decay vertices cannot be further distinguished from hadronic interactions, the true decay vertices within the vetoed regions are removed as well, thus reducing the overall reconstruction efficiency. Hadronic interaction rejection windows around peaks in the radial distribution of two-track vertices are chosen to balance the elimination of hadronic interaction vertices against the degradation of the true decay vertex reconstruction efficiency.

3.3 Single secondary vertex search

All tracks in a jet contributing to the cleaned set of two-track vertices are then combined into one list of selected tracks. This track list is supplied to the vertex fitter. The vertex fitter runs iteratively on all the tracks in the list, trying to fit one secondary vertex from all these tracks. In each iteration the track with the largest χ^2 of the track-vertex association is removed and the vertex fit is repeated until an acceptable vertex χ^2 and a vertex invariant mass < 6 GeV are obtained.

If the obtained vertex after the iterative fit procedure contains only two tracks, the vertex cleaning (see Section 3.2) is applied again; this step reduces considerably the fake rate of the SSVF.

4 Secondary vertex based b -tagging algorithms

The secondary vertex b -tagging algorithms take the information of the secondary vertices provided by the SSVF algorithm described in this note and use it to construct discriminating variables for b -tagging jets. The tagging algorithms based on the identification of a single vertex, the SV x family [2, 7], are classified into SV0, SV1, or SV2, according to the information used to calculate the discriminating variables.

The SV0 algorithm uses the significance of the 3D distance between secondary and primary event vertices. For each jet, the SV1,2 tagging algorithms calculate the ratio of the multivariate b -jet vertex properties probability density function to the light-jet probability density function. This likelihood ratio is used as a discriminating variable. The SV1,2 likelihood ratios can be easily combined with other likelihood ratios, calculated by e.g. track impact parameter based b -tagging algorithm, to increase the b -tagging performance [3].

4.1 SV0 b -tagging algorithm

The SV0 jet tagging algorithm [8, 9] uses the 3D signed decay length significance as a discriminating variable. The decay length (L) is the three-dimensional distance between the reconstructed SV and the primary vertex position. The decay length significance is defined as the ratio of the decay length to its uncertainty L/σ_L . The sign is fixed by the projection of the decay length vector (the vector pointing from the primary event vertex to the secondary vertex) onto the jet axis. The SV0 algorithm requires the signed 3D decay length significance to be positive and above some cut value to identify a b -jet.

4.2 SV1/SV2 b -tagging algorithms

The more powerful SV1 and SV2 algorithms use a likelihood ratio test to discriminate between light and b -jets. The corresponding probability density functions (PDF) are constructed using four variables:

- The total invariant mass, m , of all particles in the reconstructed secondary vertex.
- The ratio, f , of sum of energies of tracks in the SV to the energy of all tracks in the jet.
- The number, n , of two-track vertices after removing those identified as corresponding to V^0 s or material interactions.
- The ΔR between the SV-PV direction and the jet direction (SV1 only). The SV-PV vector is originating from the PV and points towards the SV.

Joint distributions of these variables are estimated using Monte Carlo (MC) simulation. To guarantee a reasonably fast convergence and stability, care is taken when defining multidimensional PDF histograms. Regions of the histograms with low predicted event density usually require large MC samples for reliable PDF estimation. This problem can be avoided by transformations of the PDF variables to make the required PDF more uniformly distributed.

The following transformations are used in the SV1/SV2 algorithms:

- SV mass (in GeV) transformation: $m' = \frac{m}{m+1}$. The m' is fully contained in the range $[0, 1]$.
- Energy ratio transformation: $f' = f^{0.7}$.

- Transformed number of two-track vertices: $n' = \log n$

The m, f transformations are chosen such that the light jet m', f' distributions occupy the lower part of the available range, while the b -jet ones occupy the upper part of the range, see Ref. [2] for details. Also the m', f' and n' distributions are more uniform than the original m, f and n ones and, therefore, are easier to sample using simulated events.

The PDF for light and b -jets are obtained by using MC simulations to fill the multidimensional histograms (2D or 3D) of the transformed variables. The obtained histograms are smoothed by the ASH method (Average Shifted Histograms) [10] and are used to estimate the corresponding probability density functions. The b -tagging likelihood ratio test should give a reasonable answer even for jets where a secondary vertex is not found. This can be achieved by adding to the b -tagging PDF a δ -function describing the jets without SVs. For these jets the variables are set to $m', f', n' = 0$ by convention. The complete jet PDF is written as

$$\text{PDF}_i = (1 - \varepsilon_i)\delta(m', f', n') + \varepsilon_i \text{ASH}_i(m', f', n'), \quad (1)$$

where ε_i is the probability to reconstruct a SV in the jet and $i = b, \text{light}$. The log of the likelihood ratio test statistic becomes for e.g. a b -jet against light jet test:

$$w = \begin{cases} \ln \left[\frac{\varepsilon_b}{\varepsilon_{\text{light}}} \cdot \frac{\text{ASH}_b(m', f', n')}{\text{ASH}_{\text{light}}(m', f', n')} \right] & \text{SV is present,} \\ \ln \left[\frac{(1 - \varepsilon_b)}{(1 - \varepsilon_{\text{light}})} \right] & \text{SV is absent.} \end{cases} \quad (2)$$

This test statistics w is usually called a b -tagging discriminating variable or a b -tagging jet weight.

The SV1 algorithm uses a direct product $2D \times 1D \times 1D$ PDF: $\text{PDF}_i(m', f') \times \text{PDF}_i(n') \times \text{PDF}_i(\Delta R)$, here ΔR is an angular distance between $\vec{SV} - \vec{PV}$ vector and the jet axis. The SV2 algorithm uses a 3D $\text{PDF}_i(m', f', n')$. For the SV1 algorithm 2D and 1D histograms are filled independently, while for the SV2 algorithm a 3D histogram is filled, which demands much larger samples. The SV1 b -tagging algorithm is simpler but expected to give a poorer rejection than the SV2 algorithm. The SV1 b -tagging algorithm is used as default in ATLAS. In this note only the SV1 algorithm performance is presented, since the SV2 one has similar features.

5 Performance

5.1 Samples

The results in this note are obtained using $t\bar{t}$ simulated events corresponding to a centre-of-mass energy of $\sqrt{s} = 13$ TeV proton-proton collisions, generated using the Powheg generator [11] interfaced to Pythia 6 [12] with the Perugia 2012 tune. EvtGen [13] is used to model the decays of b - and c -hadrons. The CT10 [14] parton distribution function set is used in the simulation. More details are available in Ref. [15]. Additional minimum-bias pp interactions are simulated with the Pythia 8 generator [16] and are overlaid on the simulated events with an average of 20 pileup interactions per event to emulate the data-taking conditions of Run-2. Only $t\bar{t}$ decays with at least one lepton from a subsequent W -boson decay are

included in the simulation. The propagation of particles through the detector and the detector response are modelled with the detailed ATLAS detector simulation [17] based on Geant4 [18].

5.2 Selection of jets

Jets are reconstructed using the anti- k_t algorithm [19]. Jets are required to have a transverse momentum of $p_T > 20$ GeV and a pseudorapidity of $|\eta| < 2.5$. In addition, jets with a transverse momentum of $p_T < 60$ GeV and pseudorapidity of $|\eta| < 2.4$ are required to pass the Jet Vertex Tagger requirement, aimed at reducing the impact of pile-up [20]. The jets are flavour labeled by matching them to weakly decaying b - and c -hadrons in the event generator record. If a b -hadron is found within a distance of $\Delta R < 0.3$ from the jet axis, then the jet is labeled as a b -jet. If no b -hadron is found, the search is repeated for c -hadrons, if a c -hadron is found and no τ leptons are found, the jet is then labeled as a c -jet. If no match is found for c , b , or τ , the jet is labeled as a light-flavour jet.

5.3 Two-track vertices cleaning performance

As mentioned in Section 3.2, some of the reconstructed two-track vertices come from K^0 , Λ^0 decays, photon conversions and interactions with the detector material. For a proper reconstruction of the b/c -decays these nuisance vertices have to be removed.

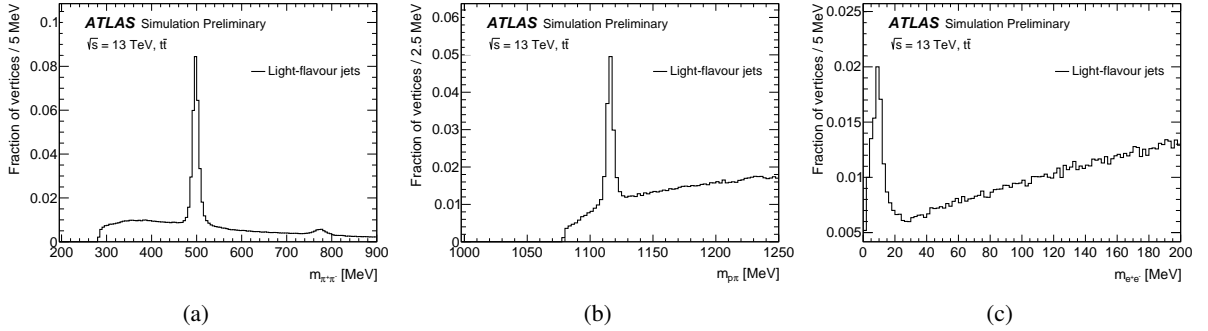


Figure 1: The invariant mass spectra of track pairs from the reconstructed vertices in light jets using the $\pi^+\pi^-$ (a), $p\pi$ (b) and e^+e^- (c) mass hypotheses. The plots are obtained using a Powheg+Pythia6 $t\bar{t}$ sample and are normalized to one, including over- and underflow bins.

Figure 1 shows the invariant mass distributions of track pairs from the reconstructed two-track vertices for the $\pi^+\pi^-$ mass hypothesis with a peak corresponding to K^0 decays, the $p\pi$ mass hypothesis with a peak corresponding to Λ^0 decays and the e^+e^- mass hypothesis with a low mass peak due to photon conversions. A second mass peak from the $\rho(770)$ decay is present in the $\pi^+\pi^-$ mass distribution as well. This peak is highly suppressed by the 3D impact parameter cut used for the track selection, see Section 3. The mass distributions are obtained by the secondary vertex finding algorithm applied to light jets in simulated $t\bar{t}$ events. The V^0 decay vertices are subsequently eliminated by requirements in mass windows around the corresponding nominal masses: $|m_{\pi^+\pi^-} - m_{K^0}| > 22$ MeV, $|m_{p\pi} - m_{\Lambda^0}| > 8$ MeV and $m_{e^+e^-} > 40$ MeV.

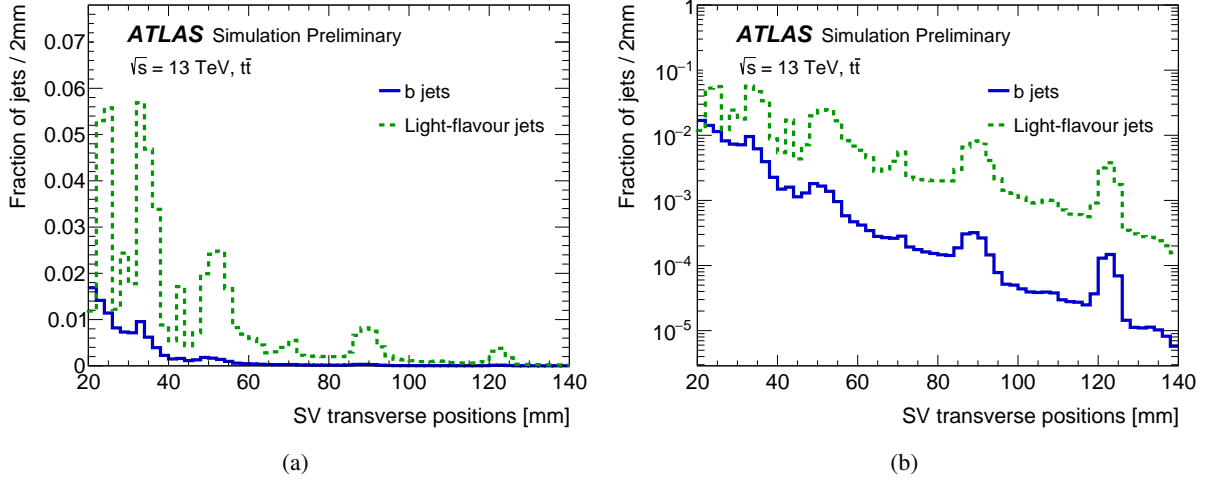


Figure 2: The transverse position of the secondary vertices in b - and light-flavour jets with peaks due to hadronic interactions in the Pixel detector layers and beampipe (a). Panel (b) shows the same distribution in a log scale for better visibility of the peaks. All distributions are normalized to one, including over- and underflow bins.

Figure 2 demonstrates the radial distribution of the two-track vertices reconstructed in light jets. Peaks due to hadronic interactions are clearly visible at radii corresponding to the Pixel detector layers (~ 34 mm, ~ 51 mm, ~ 89 mm, ~ 123 mm) and beampipe (~ 25 mm)². The distributions in Figure 2 are normalized to one in the shown range. A larger peak close to 0 mm, which originates from fake vertices from the PV, is not shown.

Figure 3 shows the cleaning effect of applying the material interaction removal procedure and the V^0 cleaning procedure, as described in Section 3.2: applying both procedures, and applying each of them separately. The figure shows the radial distance between the final reconstructed secondary vertices and the nearest material layer. Hadronic interaction vertices are concentrated at the low end of these distributions, as can be clearly seen in the light jets plot, where a contribution of such vertices is significant. As expected, the vertices in b -jets are practically unaffected by the combined vertex cleaning procedure, while the number of reconstructed vertices in the light jets is significantly reduced.

After application of the cleaning procedure the remaining reconstructed vertices in light jets are mainly fake vertices produced by occasional intersections of random tracks. The real detector material structure is much more complex than a simple geometrical model used in the cleaning procedure, therefore this procedure is not able to remove all interaction vertices. Leftovers of the cleaning procedure constitute another part of light jet vertices. Other sources of secondary vertices in jets, like the real displaced $K \rightarrow 3\pi$ decay vertices, give a negligible contribution.

² More information about the Pixel detector material and hadronic interactions in it can be found in Ref. [21].

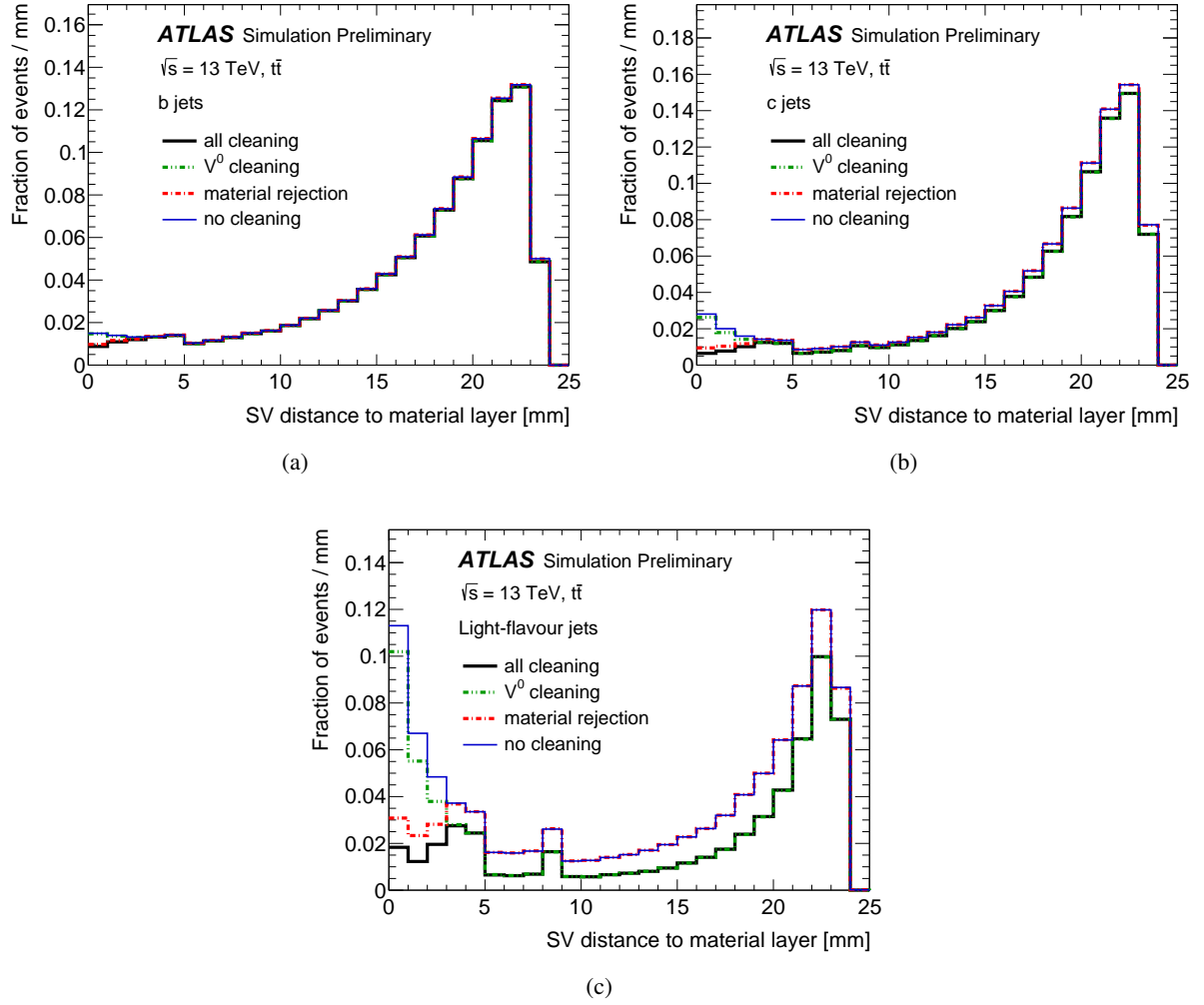
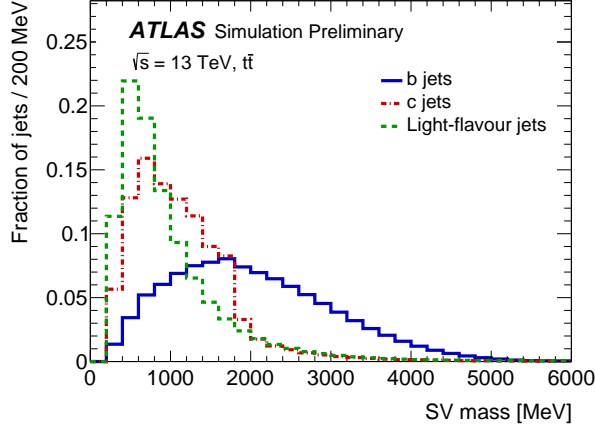


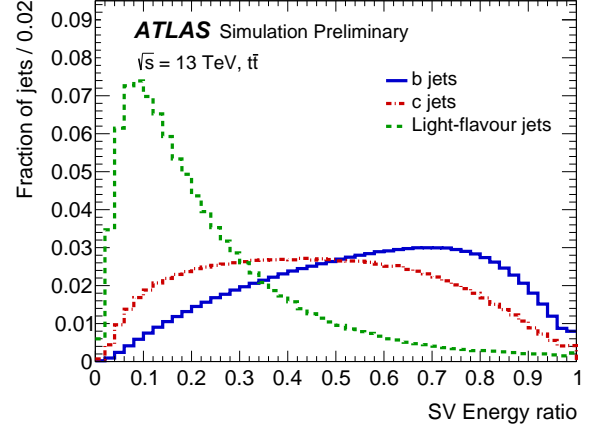
Figure 3: The distribution of the distance between the SV and the nearest material layer for different selection cuts for (a) b -jets, (b) c -jets, (c) light jets. Both of the cleaning algorithms are applied (black), no cleaning is applied (blue), only material rejection is applied (red), only V^0 cleaning is applied (green). The distributions are normalized to the curve where no cleaning is applied.

5.4 Properties of secondary vertices

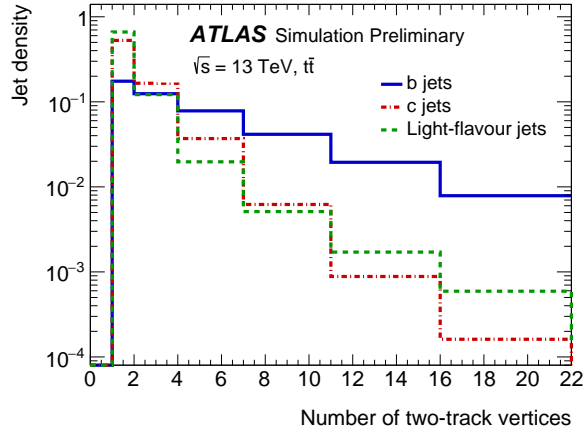
Several variables are computed by the Secondary Vertex Finder algorithm for the final secondary vertices. The normalized distributions of the SV properties are shown in Figure 4: the SV mass, the energy ratio of the tracks contributing to the SV to the sum of energies of all the tracks in the jet, the number of two-track vertices, ΔR of the flight direction with respect to the jet axis, the SV transverse distance to the primary vertex and the 3D SV-PV distance divided by its error (3D decay length significance). The SV transverse position plot shows the radial distribution of reconstructed vertices after the cleaning procedure. Not all hadronic interaction vertices can be removed without reducing the efficiency for real vertices, some enhancements around the material layers positions remain. Some of those variables are used as inputs to the SV based b -tagging algorithms. Figures 5 and 6 show the profile distributions (the mean value) of these input variables as a function of jet η and jet p_T . These variables were selected due to their power to discriminate the b -/ c -hadron decay vertices from those in light jets and relatively weak dependence as a function of jet η . However, the number of two-track vertices reaches a maximum value in the central η region inside b -jets. Also the $\Delta R(\text{Jet}, \text{SV-PV})$ distribution, where SV-PV is the direction from the primary vertex to the SV, is not flat for light jets. On the other hand, the reconstructed secondary vertex mass, the energy fraction and the SV distance to material layers are independent of jet η . The mean values of the same variables are also shown as a function of jet p_T in the same Figures 5 and 6. The distributions illustrate one of the challenges faced by SV reconstruction in high- p_T jets. The higher p_T jets have, the higher is the track multiplicity and track density at the jet core, which causes degradation of the track reconstruction efficiency. The degradation in the quality of reconstructed secondary vertices is due to decreasing the amount of reconstructed tracks from the b - or c -hadron decays, and also due to the high track density which leads to the increase of the amount of fake two-track vertices. This effect rises significantly at high jet p_T , and the SV properties are becoming less discriminating as shown in the figures.



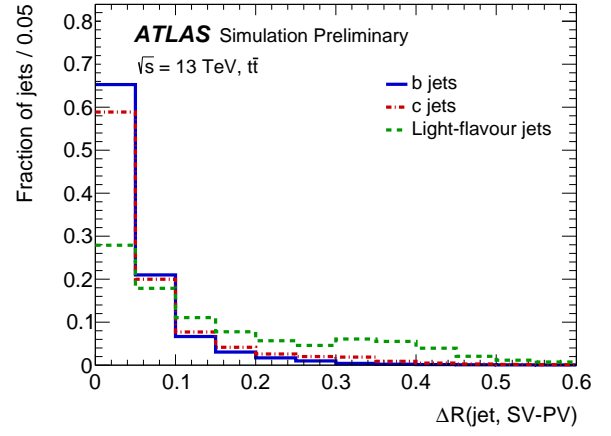
(a)



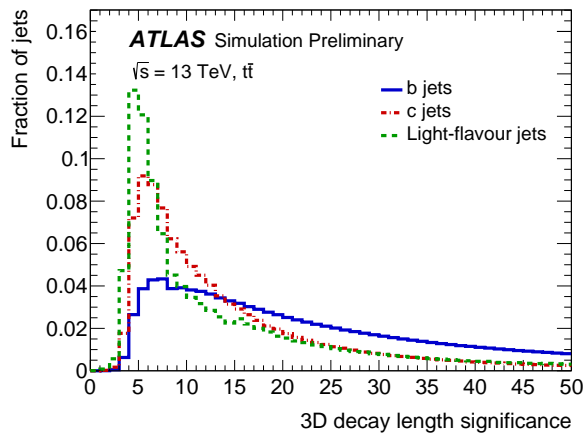
(b)



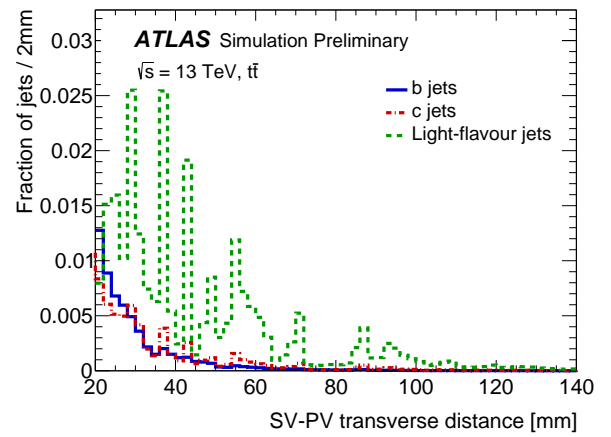
(c)



(d)

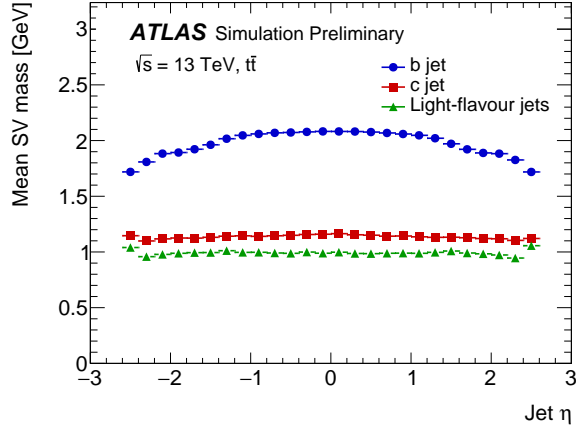


(e)

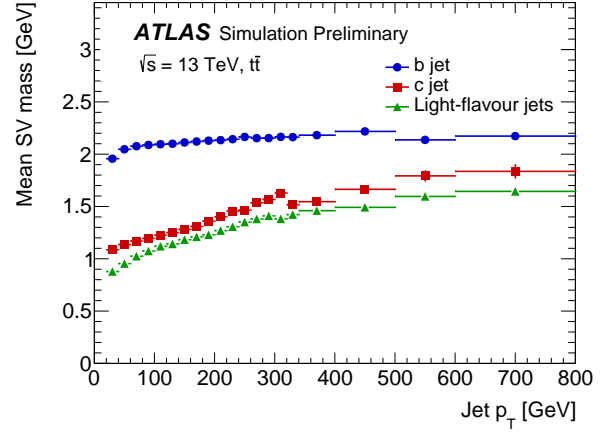


(f)

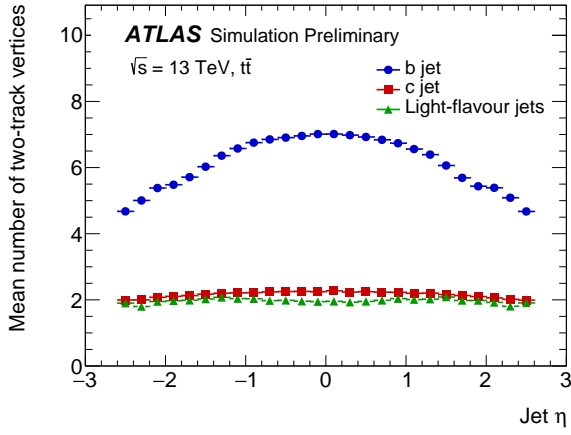
Figure 4: Secondary vertex properties distributions with b -jet (blue), c -jet (red) and light jet (green). All distributions have been normalized to one, including over- and underflow bins.



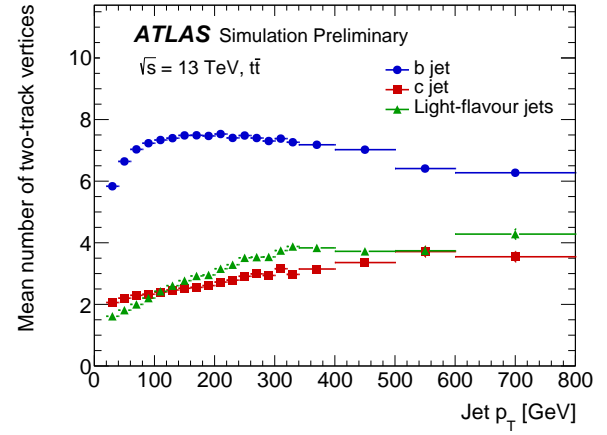
(a)



(b)

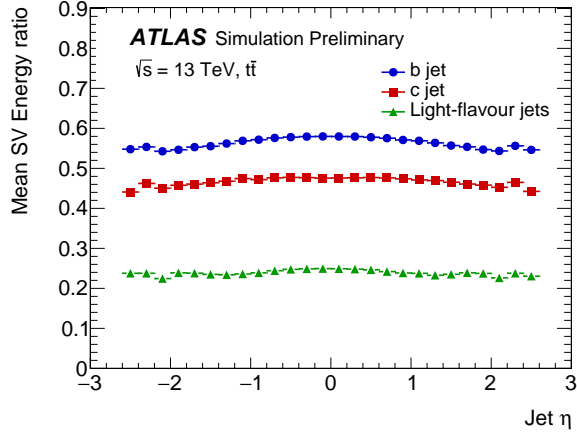


(c)

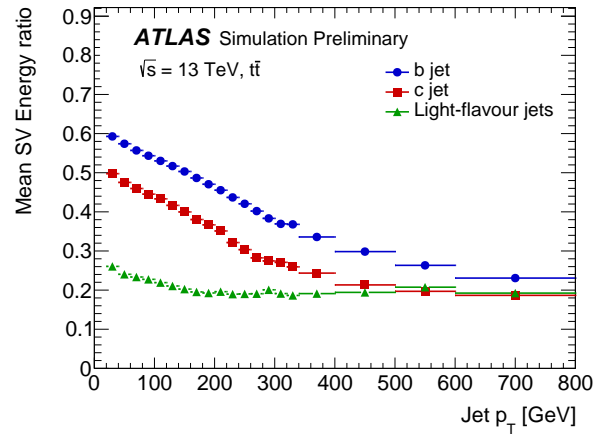


(d)

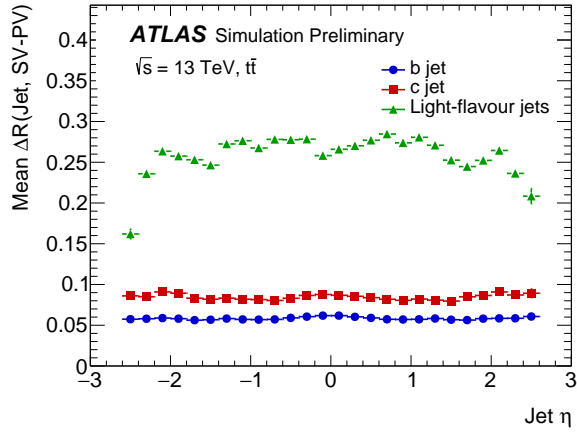
Figure 5: The mean values of secondary vertex properties as a function of jet η (left column) and jet p_T (right column) with b -jet(blue), c -jets (red), light jets (green).



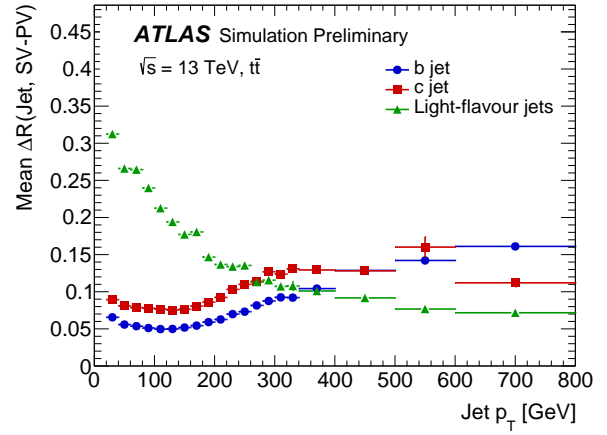
(a)



(b)



(c)



(d)

Figure 6: The mean values of further secondary vertex properties as a function of jet η (left column) and jet p_T (right column) with b -jet(blue), c -jets (red), light jets (green).

5.5 Performance of the secondary vertex finder

The secondary vertex reconstruction efficiencies, i.e. the fractions of jets which have a reconstructed secondary vertex, for different jet flavours of the SVF are plotted as a function of jet η in Figure 7. An average reconstruction efficiency of $\sim 80\%$ is found in b -jets, with a drop for $|\eta| \geq 1.5$ being observed for both b - and c -jets. The probability to have fake secondary vertices in light jets shows very low values in the central region but an increase at high $|\eta|$ regions. Both effects can be explained by the larger amount of detector material in the endcap region for $|\eta| > 1.5$, resulting in higher hadronic interaction rate and decrease of the track reconstruction efficiency, and decrease of spatial accuracy of the track reconstruction there. In addition, the degraded tracking resolution augments the pileup contribution to the fake secondary vertex rate in the endcap region.

The SV reconstruction efficiencies for different jet flavours as a function of jet p_T are shown in Figure 7. The increase of track density in high- p_T jets leads to degradation of the track reconstruction efficiency and consequently smaller amount of reconstructed tracks from b/c -hadron decays [6, 22]. A larger amount of fragmentation tracks in a jet is responsible for the larger fake rate of reconstructed vertices, as well. As shown in the figure, the number of fake vertices is increasing with jet p_T in light jets, while the SV reconstruction efficiencies for b -jets slightly degrade with jet p_T .

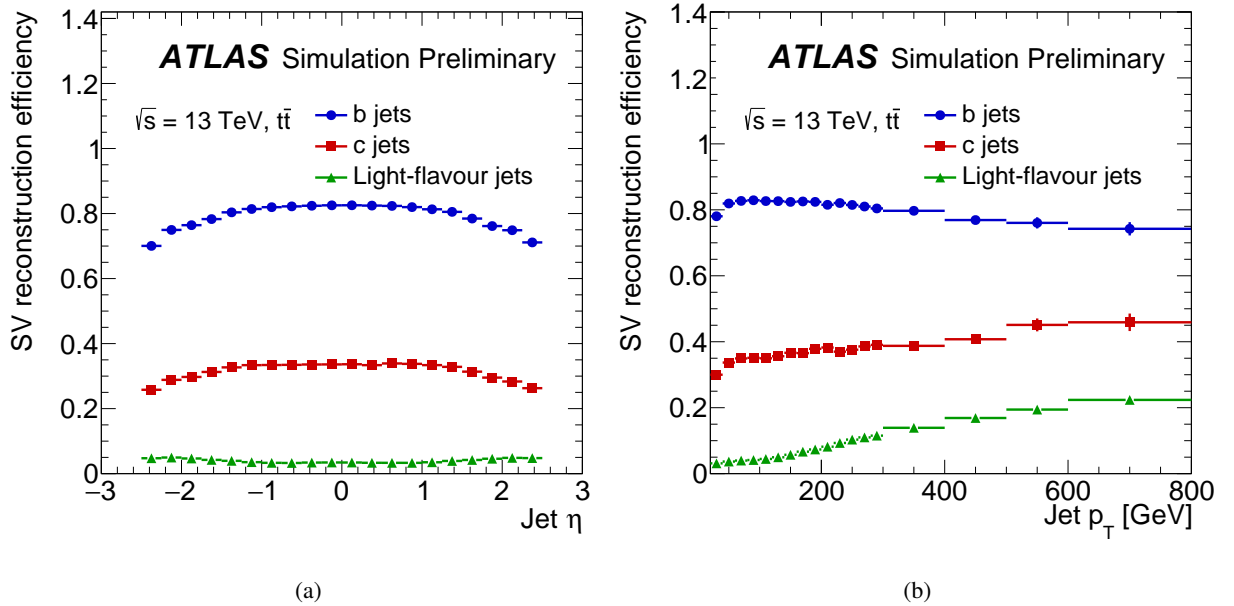


Figure 7: The secondary vertex reconstruction efficiencies for three jet flavours, as a function of jet η (a), as a function of jet p_T (b). b -jets in blue, c -jets in red, and fake vertices are shown in green for light jets.

5.6 Performance of the secondary vertex based b -tagging algorithms

The performance of a secondary vertex based b -tagging algorithm is limited by the ability to reconstruct secondary vertices in a b -jet. As shown in previous sections (Figure 5), the SV reconstruction efficiency in a b -jet reaches a maximum $\sim 80\%$ at the jet transverse momentum ~ 80 GeV, then it gradually decreases at higher p_T . A more universal and powerful b -jet tagging algorithm can be constructed by combining the SV based algorithm with the track impact parameter based algorithm IP3D [3]. When a SV is reconstructed in a jet, a combination of the tracking information from IP3D and kinematical information from SV provides a significant improvement of the light jet rejection. When a SV is not reconstructed in a jet, the combined algorithm is still able to identify the b -jet using the IP3D information only, although with higher fake rate. The SV based algorithms and the impact parameter based algorithm IP3D provide jet weights depending on the jet flavour content. These weights are logarithms of the likelihood ratios and therefore can be combined by summing them, corresponding to the multiplication of the corresponding likelihoods.

The light and c -jet rejections versus the b -tagging efficiency for the algorithms SV0, SV1 and the combination of SV1 and IP3D are shown in Figure 8. The b -tagging efficiency is defined as the fraction of jets that are labeled as b -jets and tagged as b -jets by a tagging algorithm, while the light rejection is the inverse of the mis-tagging rate: The mis-tagging rate is the fraction of jets labeled as light jets and tagged as b -jets. The SV0 and SV1 p_T and η distributions as a function of the light jet rejection are shown in Figure 9 for a b -tagging efficiency of 70%, as determined in $t\bar{t}$ simulated events.

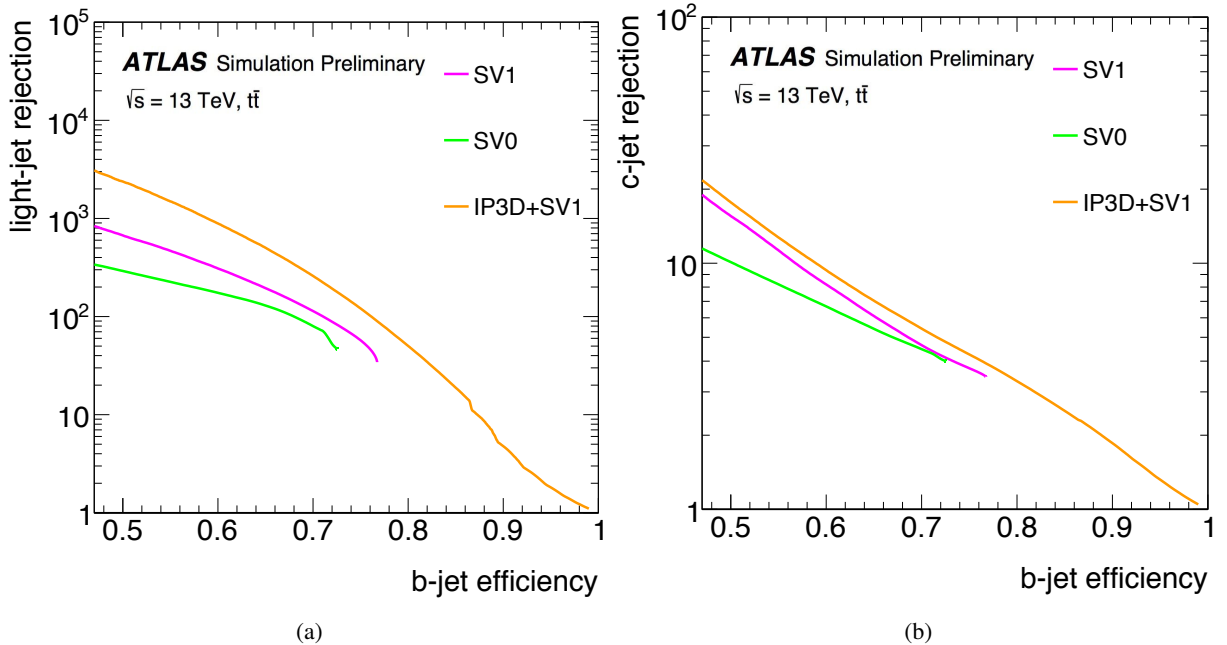


Figure 8: Performance of the SV0, SV1 and SV1+IP3D b -tagging algorithms. Light jet rejection (a) and c -jet rejection (b) versus b -jet efficiency.

However, this simple approach of combining jet tagging algorithms neglects correlations between them. A better handling of the information provided by the individual algorithms can be achieved by exploiting multivariate approaches such as neural networks or boosted decision trees (BDT). The current baseline

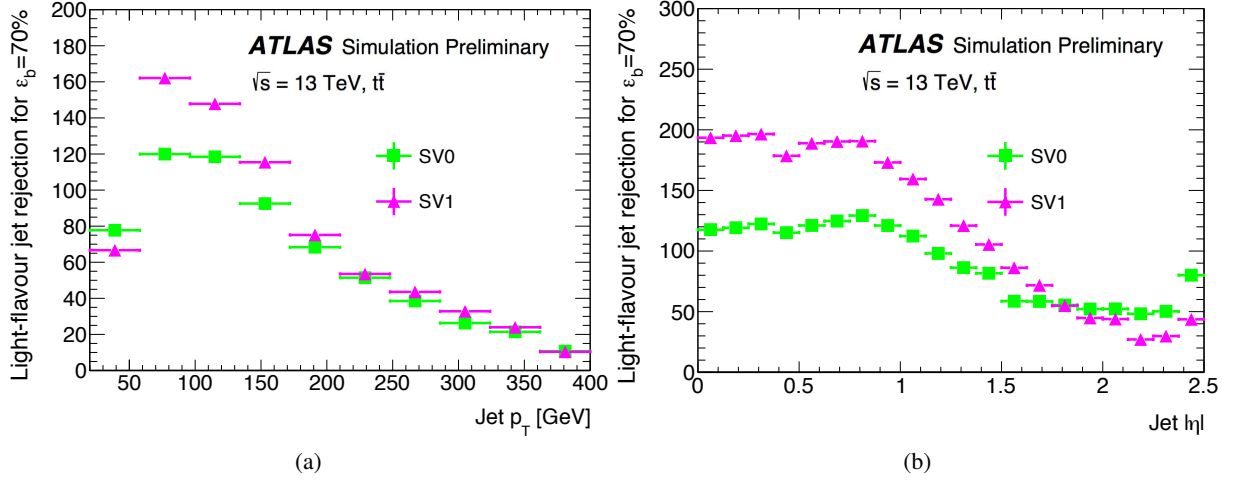


Figure 9: The SV0 and SV1 light jet rejection for a b -tagging efficiency of 70% as a function of the jet p_T (a) and η (b).

ATLAS b -tagging algorithm MV2 [5, 6] uses kinematic and topological information provided by the SV finder, together with information provided by several other algorithms, as input to a dedicated BDT algorithm.

6 Conclusion

The note presents a unified approach to exploit the presence and the kinematic features of secondary vertices in a jet for the jet flavour identification. A method for vertex finding in a given set of tracks is introduced. To identify and reconstruct secondary vertices in the jets the Single Secondary Vertex Finder implementation of the method has been developed, which shows a high performance and has a low fake rate. The kinematic features of the secondary vertices, highly sensitive to the flavour content of the jets, are used to construct several high performance b -jet tagging algorithms which can be used either separately or in combination with other b -tagging algorithms.

The SVF algorithm is optimized and tuned to the information provided by the ATLAS detector to achieve the highest reconstruction efficiency and purity of the secondary vertices. In the central region of the detector a typical SV reconstruction efficiency for b -jets of 80% is achieved, with c -jet and light-jet efficiencies of 30% and <5%. This efficiency reaches a maximum for jets with a transverse momentum of 80 GeV.

The secondary vertex based b -tagging algorithm efficiencies and rejections have been presented in detail. The performance of the SV0, SV1 and of the combined IP3D+SV1 taggers are compared. For a b -tagging efficiency working point of 50%, the light-jet rejection improves from 290 to 670 and 2400, while the c -jet rejection is 10, 16 and 18, respectively, for the three taggers.

References

- [1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, **JINST** **3** (2008) S08003.
- [2] V. Kostyukhin, *Secondary vertex based b -tagging*, ATL-PHYS-2003-033, 2003, URL: <http://inspirehep.net/record/1195848>.
- [3] ATLAS Collaboration, *Performance of b -Jet Identification in the ATLAS Experiment*, **JINST** **11** (2016) P04008, arXiv:1512.01094 [hep-ex].
- [4] M Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, ATLAS-TDR-19, 2010, URL: <https://cds.cern.ch/record/1291633>.
- [5] ATLAS Collaboration, *Expected performance of the ATLAS b -tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, 2015, URL: <https://cds.cern.ch/record/2037697>.
- [6] ATLAS Collaboration, *Optimisation of the ATLAS b -tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: <https://cds.cern.ch/record/2160731>.
- [7] ATLAS Collaboration, *Vertexing for b -Tagging*, ATL-PHYS-PUB-2009-019, 2009, URL: <https://cds.cern.ch/record/1174260>.
- [8] ATLAS Collaboration, *Performance of the ATLAS Secondary Vertex b -tagging Algorithm in 900 GeV Collision Data*, ATLAS-CONF-2010-004, 2010, URL: <https://cds.cern.ch/record/1273194>.
- [9] ATLAS Collaboration, *Performance of the ATLAS Secondary Vertex b -tagging Algorithm in 7 TeV Collision Data*, ATLAS-CONF-2010-042, 2010, URL: <https://cds.cern.ch/record/1277682>.
- [10] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*, John Wiley & Sons, 2015.
- [11] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, **JHEP** **11** (2004) 040, arXiv:hep-ph/0409146.
- [12] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, **JHEP** **05** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].
- [13] D. Lange, *The EvtGen particle decay simulation package*, **Nucl. Instrum. Meth.** **A462** (2001) 152.
- [14] H.-L. Lai et al., *New parton distributions for collider physics*, **Phys. Rev.** **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].
- [15] ATLAS Collaboration, *Simulation of top-quark production for the ATLAS experiment at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2016-004, 2016, URL: <https://cds.cern.ch/record/2120417>.
- [16] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, **Comput. Phys. Commun.** **178** (2008) 852, arXiv:0710.3820 [hep-ph].
- [17] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, **Eur. Phys. J.** **C70** (2010) 823, arXiv:1005.4568 [hep-ex].
- [18] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, **Nucl. Instrum. Meth.** **A506** (2003) 250.

- [19] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, **JHEP** **04** (2008) 063, arXiv:[0802.1189](#) [[hep-ph](#)].
- [20] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, **Eur. Phys. J. C** **76** (2016) 581, arXiv:[1510.03823](#) [[hep-ex](#)].
- [21] ATLAS Collaboration, *Studies of the ATLAS Inner Detector material using $\sqrt{s}=13$ TeV pp collision data*, ATL-PHYS-PUB-2015-050, 2015, URL: <https://cds.cern.ch/record/2109010>.
- [22] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC run 2*, (2017), arXiv:[1704.07983](#) [[hep-ex](#)].