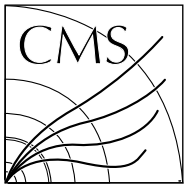
Available on CMS information server

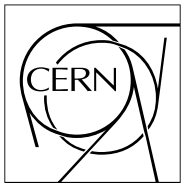
CMS NOTE 2006/014



The Compact Muon Solenoid Experiment

CMS Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland



January 25, 2006

A Combined Secondary Vertex Based B-Tagging Algorithm in CMS

C. Weiser

Institut für Experimentelle Kernphysik, Universität Karlsruhe, Germany

Abstract

This Note describes a b-tagging algorithm based on inclusive secondary vertex reconstruction in jets. Several topological and kinematical secondary vertex related variables as well as information from track impact parameters are combined into a single tagging variable to discriminate between jets originating from b-quarks and those from other sources. The performance of the algorithm is studied for several event topologies and as a function of jet kinematics. Effects of an imperfectly aligned tracking detector on the performance are also investigated.

1 Introduction

Many physics channels have b-quark jets in the final state, such as events containing top quarks, Higgs bosons or supersymmetric particles. The top quark decays almost exclusively into a W-boson and a b-quark and for low masses of the Standard Model Higgs boson ($m_H \leq 135$ GeV/c^2) its dominant decay is into a pair of b-quarks, $\mathrm{H} \rightarrow \mathrm{b\overline{b}}$. Discriminating between jets originating from b-quarks and those from other sources (d-, u-, s-, c-quarks or gluons) is crucial for many channels in order to isolate signal events from background events. In multi-jet final states, efficient b-tagging and high rejection factors against non-b jets are often needed to determine the correct jet configuration within an event (e.g. in semi-leptonic or fully hadronic $\mathrm{t\overline{t}}$ events to find out which non-b jets stem from the W decay and which b-jets from the decays of the top quarks).

Identifying b-jets relies on the properties of the production and the weak decay of b-hadrons. The most important property is the relatively large lifetime of b-hadrons of about 1.5 ps ($c\tau \approx 450\mu$m) corresponding to a flight distance that is observable with high resolution tracking detectors. This leads to secondary vertices displaced from the primary event vertex and charged particle tracks incompatible with the primary vertex. b-hadrons have a large mass and large multiplicity of charged particles in the final state (about five charged particles on average per b-hadron decay). Because of the hard b-fragmentation function, the b-hadron in a b-jet carries a large fraction of the jet energy. Since b- and c-hadrons may decay semi-leptonically, in about 20% (per lepton species) of the cases an electron or muon is produced inside a b-jet, if both direct and cascade decays are taken into account.

This Note describes a b-tagging algorithm based on inclusive secondary vertex reconstruction in jets. It combines several kinematical and topological secondary vertex variables as well as the information from track impact parameters.

The Note is organised as follows: Section 2 describes the physics input objects needed for b-tagging, Section 3 provides the definitions needed for the performance studies, Section 4 presents the combined secondary vertex based b-tagging algorithm whose performance is then discussed in Section 6. In Section 7, the robustness of the algorithm against misalignment of the tracking detectors is studied. Finally, a summary is given in Section 8.

2 Reconstructed objects used as input

Algorithms for b-tagging rely on the reconstruction of basic physics objects. At the LHC, b-tagging will be typically applied to jets. Powerful jet reconstruction algorithms are thus necessary. Several jet reconstruction algorithms are available in CMS. They are described together with their performance in [1]. For the studies described in this Note, the iterative cone algorithm with a cone size of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.5$ was used. The input used for the jet clustering are the towers from the electromagnetic and hadronic calorimeters applying a variable noise subtraction. A calibration as deduced from the Monte Carlo simulation has been applied to correct the raw jet energy.

Most of the b-hadron properties used for b-tagging can only be exploited using charged particle tracks because only tracking detectors offer the spatial resolution needed to resolve properties of b-hadron decays such as their significant flight path. Efficient track reconstruction, and in particular precise spatial reconstruction close to the interaction point, is thus the key ingredient. Track finding was performed using a combinatorial track finder applying a Kalman filter technique [2]. For b-tagging, a lifetime-based definition of the sign of the track impact parameters is used. The impact parameter is signed as positive if the track is reconstructed downstream of the primary vertex with respect to the jet direction, negative otherwise.

The following track selection cuts are applied:

- at least 8 reconstructed hits in total (pixel and silicon strip detectors);
- at least 2 reconstructed hits in the pixel detectors;
- transverse momentum $p_t > 1$ GeV/c;
- χ^2/dof of the track fit < 10;
- transverse impact parameter with respect to the reconstructed primary vertex < 2 mm to reject charged particle tracks having their origin from sources showing much larger displacement from the primary vertex (e.g. V^0 decays, photon conversions and nuclear interactions in the beampipe or the first layers of the pixel detector); to first order, the impact parameter is invariant under boosts of the b-hadron.

Charged particle tracks are associated to a jet if they are within a cone of $\Delta R < 0.3$ measured with respect to the jet axis.

To measure a flight path or displaced tracks not having their origin at the primary event vertex, precise reconstruction of the coordinates of the primary vertex is crucial. For the studies described in this Note, the primary vertex reconstructed globally in the event is used [3]. The efficiency to find the correct primary vertex and the position resolution depend on the event topology (e.g. the number of charged particles originating from the primary vertex). The position resolution is typically between 10-40 μm in the transverse plane and 15-50 μm in the z-direction. The primary vertex finding efficiency is greater than 95% for most of the physics channels with a sufficiently large multiplicity of charged particle tracks in the final state (see [3] for details).

3 Definitions

The efficiency, ϵ_q, to tag a jet of a certain flavour q as a b-jet (b-tagging efficiency for b-jets, misidentification efficiency for non-b jets) is defined as:

$$\epsilon_q = \frac{\text{Number of jets of flavour q tagged as b}}{\text{Number of jets of flavour q}}. \tag{1}$$

In the simulation studies the true flavour of a reconstructed jet was determined by analysing the parton content in a cone around the jet direction, where the assignment of a parton flavour to the jet follows a "physics" based definition. A reconstructed jet is considered to be matched to the parton produced in the primary process if it is within a cone of radius $\Delta R < 0.3$ (in QCD events the initial partons are those produced in the hard interaction, in $\mathrm{t\bar{t}}$ events the initial partons are the quarks from the top decays and hadronic W decays). Gluon jets and quark jets with b- or c-quarks from gluon splitting among their shower products thus get the flavour of the gluon or quark. To ensure an unambiguous assignment, jets are rejected if more than one initial parton fulfils this requirement.

One can also apply an "algorithmic" definition by analysing the final state partons after splitting and radiation and assigning the parton flavour that most likely determines the structure of the jet. Here, a jet originating from a gluon that has split into a $\mathrm{b\bar{b}}$ pair would get the flavour of the b-quark assigned, because from the algorithm point of view the jet looks in many aspects as a genuine b-jet. The main difference in these definitions is for gluon jets, where the non-negligible splitting rates into $\mathrm{b\bar{b}}$ and $\mathrm{c\bar{c}}$ pairs impose a serious limitation for b-tagging (depending on the energy of the gluon, these splitting rates are typically some percent).

For these studies, samples of QCD jets were used. Jets in these samples cover a wide range of transverse momentum (from about 30 GeV/c to more than 200 GeV/c) and the full geometrical acceptance of the tracking detector in pseudorapidity ($|\eta| < 2.4$). Semi-leptonic $\mathrm{t\bar{t}}$ events have also been studied.

The b-tagging performance for specific physics channels and effects like overlapping jets in dense hadronic environments will be studied in detail in dedicated Notes.

4 Algorithm description

The combined b-tagging algorithm is based on the reconstruction of the secondary decay vertex of the weakly decaying b-hadron. A discriminating variable, able to distinguish b-quark jets from non-b jets, is built by combining different topological and kinematical variables with the track impact parameter significances.

Secondary vertices are reconstructed in an inclusive way inside the jet using the *Trimmed Kalman Vertex Finder* [3]. This algorithm begins by using all tracks in the jet and subsequently rejects outliers which then are used to reconstruct additional vertices. Since not only the presence of a secondary vertex is used in the algorithm, but also topological and kinematical variables related to the vertex, it is desirable to reconstruct as completely as possible the decay vertex in order to increase the discriminating power of these topological and kinematical variables. The performance of the inclusive secondary vertex finding in jets is described in a dedicated Note [4].

The following cuts are applied to the reconstructed vertices to select secondary vertex candidates:

- The distance between the primary vertex and the secondary vertex in the transverse plane has to exceed 100 μm and must not exceed 2.5 cm.
- The distance between the primary vertex and the secondary vertex in the transverse plane divided by its error has to be greater than 3: $\frac{L_t}{\sigma_{L_t}} > 3$.

- The invariant mass of charged particles associated to the vertex must not exceed 6.5 GeV/c^2.
- The vertex must not be compatible with a K^0_S decay. Vertices with two oppositely charged tracks are rejected if their mass is within a window of $50\ \mathrm{MeV}/c^2$ around the nominal K^0_S mass.

The secondary vertex reconstruction and selection leads to three categories defined as:

1. *RecoVertex*: At least one secondary vertex candidate is reconstructed and satisfies the selection criteria. All tracks from all accepted vertices are used for the computation of the vertex related variables if there is more than one accepted secondary vertex.
2. *PseudoVertex*: If no reconstructed secondary vertex candidate is found, a so-called PseudoVertex is created using charged particle tracks not compatible with the primary vertex, having a signed transverse impact parameter significance greater than two, if at least two such tracks are present in the jet.
3. *NoVertex*: If neither 1) or 2) are fulfilled.

The distribution of the vertex categories for the different jet flavours is shown in Fig. 1. It can be seen that the presence of a secondary vertex alone is already strongly discriminating between b-quark jets and other jets.

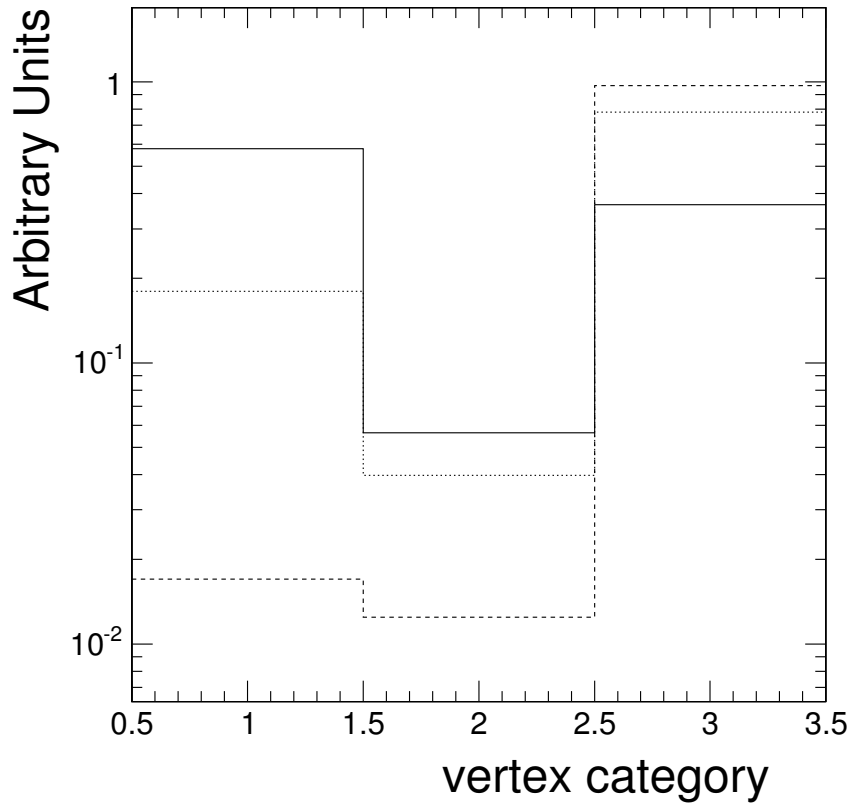


Figure 1: Distribution of the vertex category, for b-jets (solid), c-jets (dotted) and uds-jets (dashed) with transverse momenta between 50 GeV/c and 80 GeV/c.

5 Input variables and combined discriminating variable

Optimal performance is achieved by combining several topological and kinematical variables related to the secondary vertex reconstruction, as well as variables related to the impact parameter significances of charged particle tracks. The choice of variables entering into the combination depends on the vertex category.

The track impact parameter significances of accepted tracks enter into the final discriminator for all categories. Their distributions are shown for the different categories in Fig. 2. For these tracks, a requirement additional to

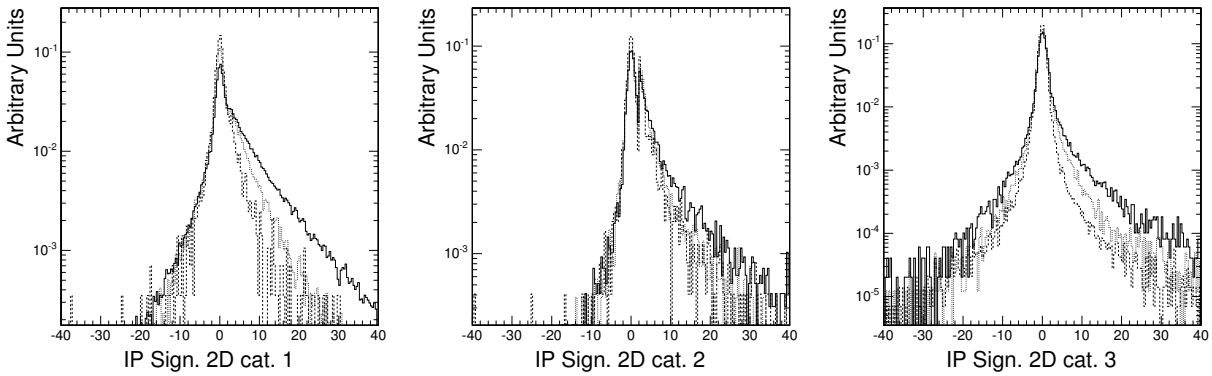


Figure 2: The distributions of the tracks impact parameter significances for category "RecoVertex" (left), "PseudoVertex" (middle) and "NoVertex" (right) for b-jets (solid), c-jets (dotted) and uds-jets (dashed).

those listed in Section 2 is requested: The distance of closest approach of the track and the jet axis is required to be smaller than 0.07 cm. Furthermore, a charged particle track is not considered for the track impact parameter variables if there is another track of opposite charge associated to the jet, that forms an invariant mass with that track within 30 MeV/c^2 around the nominal K^0_S mass.

The rejection of charm quark jets can be improved with the following procedure: Tracks are sorted in decreasing order of their impact parameter significance. The invariant mass is computed for tracks 1 to n. If this mass exceeds a value related to the mass of charm hadrons, the impact parameter significance of the track moving the n-track mass above this threshold can be added to the discriminator. The mass threshold is set to 1.5 GeV/c^2. This is lower than typical charm hadron masses because neutral particles are missed and not all charged particles from the decay are reconstructed and accepted.

Vertex category 1 ("RecoVertex"), where a reconstructed secondary vertex has been found, has the largest discriminating power. For this category, the following variables are defined:

- The invariant mass of charged particles associated to the secondary vertex. For secondary vertices in b-jets, the vertex mass can be significantly above the mass of charm hadrons, allowing to suppress efficiently this background.
- The multiplicity of charged particles associated to the secondary vertex; b-hadron decays show a significantly larger track multiplicity than e.g. charm hadron decays.
- The distance between the primary vertex and the secondary vertex in the transverse plane divided by its error. This is sensitive to the large flight path of b-hadrons.
- The energy of charged particles associated to the secondary vertex divided by the energy of all charged particles associated to the jet. This quantity is sensitive to the hard fragmentation function of b- and c-quarks.
- The rapidities of charged particle tracks associated to the secondary vertex with respect to the jet direction: $y = \frac{1}{2} \cdot \ln(\frac{E+p_\parallel}{E-p_\parallel})$. This variable enters for n tracks, where n is the charged particle multiplicity of the secondary vertex. For b-jets, the rapidities are on average smaller than for c-jets because of the large mass of b-hadrons resulting in larger angles of the charged particle tracks relative to the jet axis.
- The track impact parameter significance of the first track exceeding the charm threshold as described above.

For vertex category 2 ("PseudoVertex"), most of these variables still offer some discriminating power between b- and non-b jets. Since there is no attempt to fit the geometrical position of the "PseudoVertex", the significance of the flight distance (the third variable defined in the list above) is not used for this category, whereas the other five variables enter into the list of input variables. For vertex category 3 ("NoVertex"), no additional variables are defined.

Since there can be strong correlations between the category and the input variables (e.g. if a well reconstructed secondary vertex with large track multiplicity has been found, there is typically also a large number of tracks

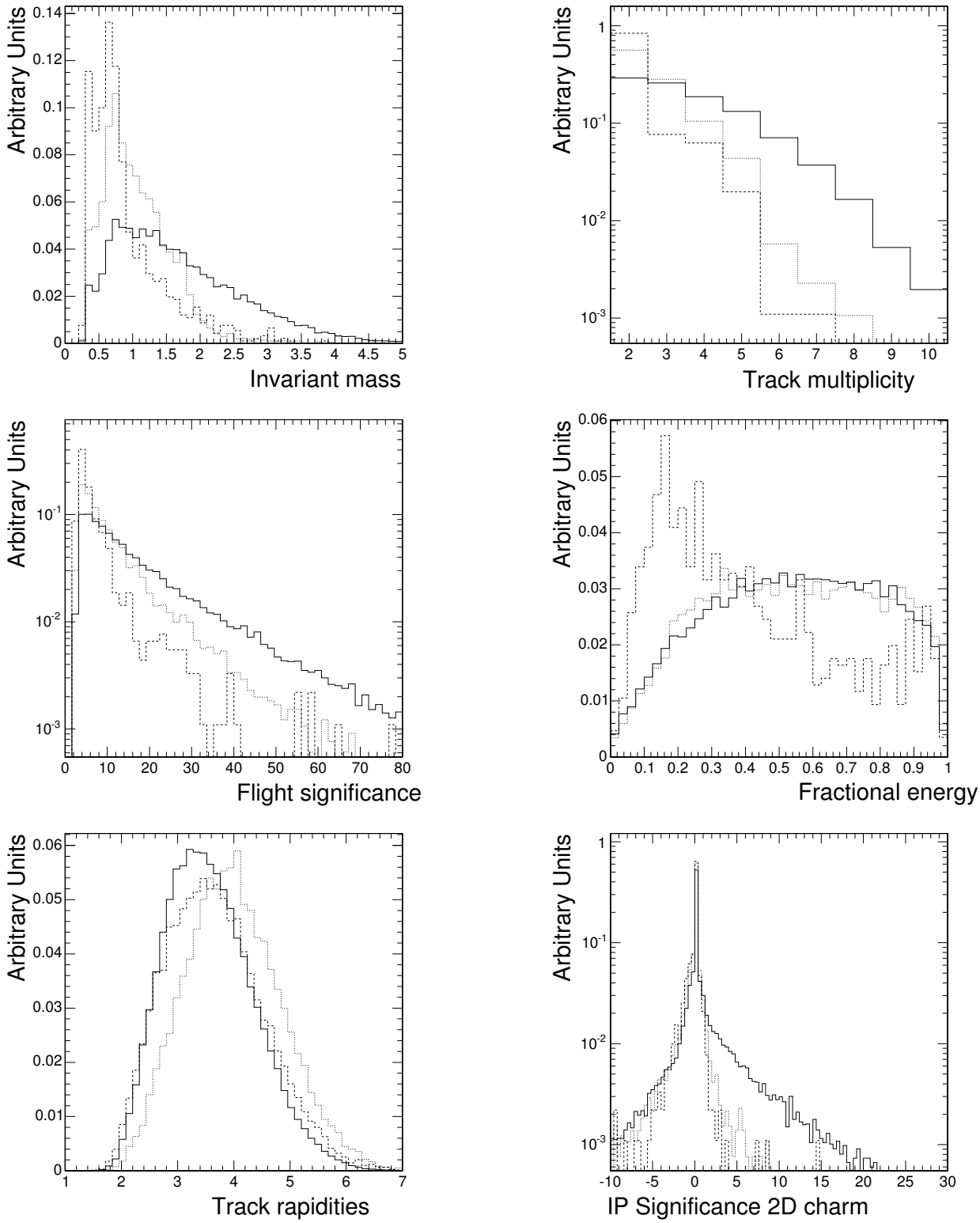


Figure 3: The distributions for the input variables for the vertex category "RecoVertex" for b-jets (solid), c-jets (dotted) and uds-jets (dashed): invariant mass (top left) and multiplicity (top right) of charged particles associated to the reconstructed secondary vertex; flight distance significance defined as the distance between the primary and secondary vertex divided by its error (middle left) and fraction of charged jet energy associated to the secondary vertex (middle right); rapidities of charged particles associated to the secondary vertex with respect to the assumed b-hadron flight direction (bottom left) and impact parameter significance of the first track exceeding the charm mass threshold as described in the text (bottom right).

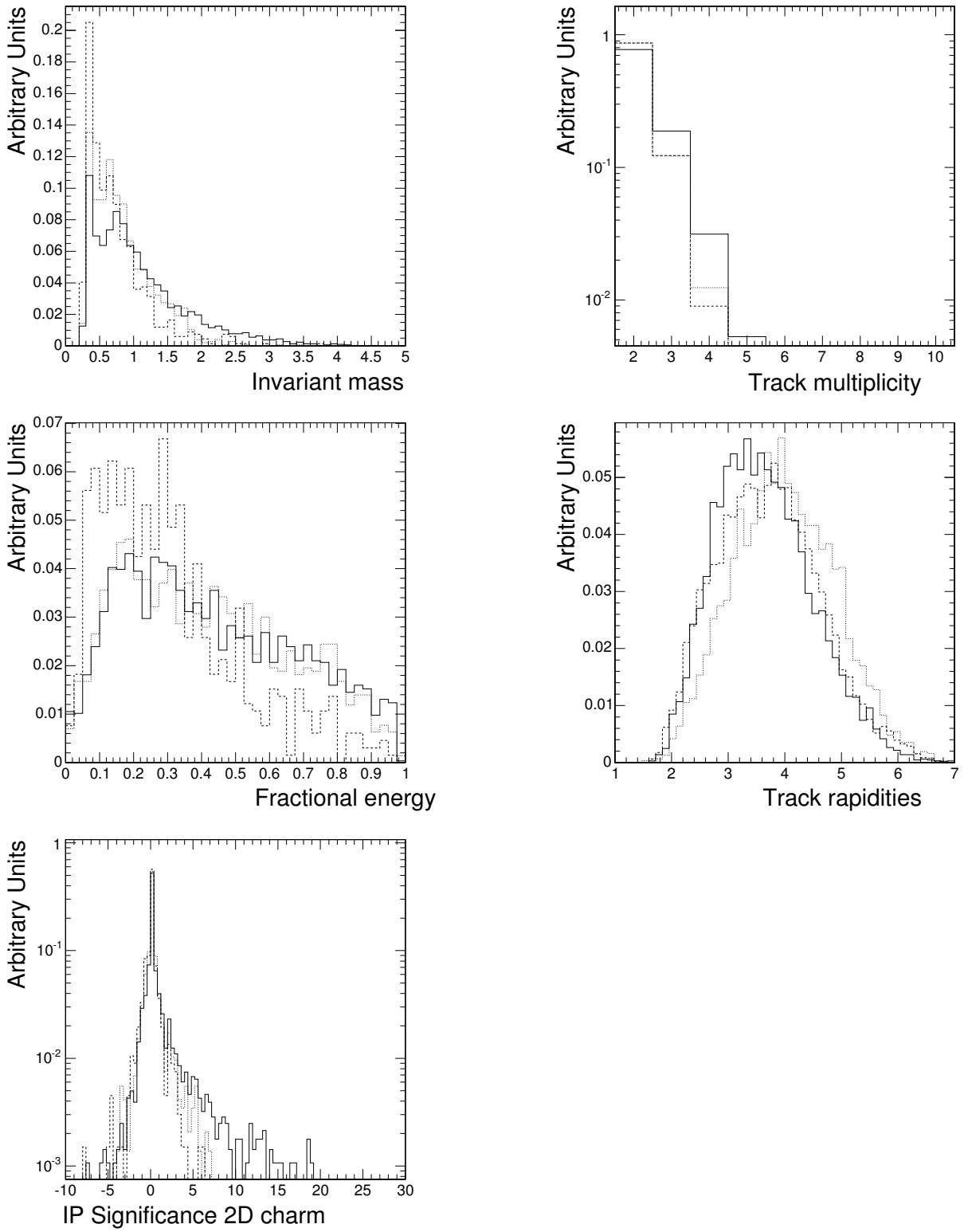


Figure 4: The distributions for the input variables for the vertex category "PseudoVertex" for b-jets (solid), c-jets (dotted) and uds-jets (dashed): invariant mass (top left) and multiplicity (top right) of charged particles associated to the secondary pseudo vertex; fraction of charged jet energy associated to the secondary vertex (middle left) and rapidities of charged particles associated to the secondary vertex with respect to the assumed b-hadron flight direction (middle right); impact parameter significance of the first track exceeding the charm mass threshold as described in the text (bottom left).

which are significantly displaced from the primary event vertex), the input variables are shown separately for the individual categories.

The input variables for category "RecoVertex" and "PseudoVertex" are shown in Figures 3 and 4, respectively.

The variables described above are combined into a single discriminating variable using a Likelihood ratio technique. Since the distributions of most of the variables look significantly different for c-jets and udsg-jets, the Likelihood ratio is split into contributions to discriminate separately the two backgrounds. The Likelihood function is defined as

$$\mathcal{L}^{b,c,q} = f^{b,c,q}(\alpha) \times \prod_i f^{b,c,q}_{\alpha}(x_i)$$

where α denotes the vertex category ($\alpha = 1, 2, 3$) as defined above, x_i are the individual variables, q stands for u,d,s-quark jets and gluon jets, $f_{BG}(c)$ and $f_{BG}(q)$ are the expected a-priori probabilities for the c- and q-content in non-b jets ($f_{BG}(c) + f_{BG}(q) = 1$), $f^{b,c,q}(\alpha)$ is the probability for flavour b,c,q to fall into category α and $f^{b,c,q}_{\alpha}(x_i)$ is the probability density function for variable x_i for category α and flavour b,c,q. $f_{BG}(c)$ and $f_{BG}(q)$ can be adapted for a specific physics process under study, if there is prior knowledge of the flavour composition of the non-b jets. For the studies presented in this Note, $f_{BG}(c) = 0.25$ (and thus $f_{BG}(q) = 1 - f_{BG}(c) = 0.75$) has been used[1)].

The combined discriminating variable, d, is defined as

$$d = f_{BG}(c) \times \frac{\mathcal{L}^b}{\mathcal{L}^b + \mathcal{L}^c} + f_{BG}(q) \times \frac{\mathcal{L}^b}{\mathcal{L}^b + \mathcal{L}^q}.$$

The probability density functions are extracted from a statistically independent sample of simulated QCD events. They depend on the transverse jet energy and pseudorapidity.

The distribution of the combined discriminating variable is shown in Fig. 5, for all categories and also separately for the categories as described above. b-quark jets have values of d close to one, whereas jets from other sources show significantly lower values on average allowing good discrimination between the different flavours. It should be noted that in these figures all flavours are normalised to unity for better visibility, whereas the population of the categories differs significantly for the individual flavours as has already been shown in Fig. 1.

[1)] This choice was made because in hadronic decays of W-bosons the expected fraction of c-jets is about 0.25. This value also represents a good compromise for other compositions of the non-b jet sample.

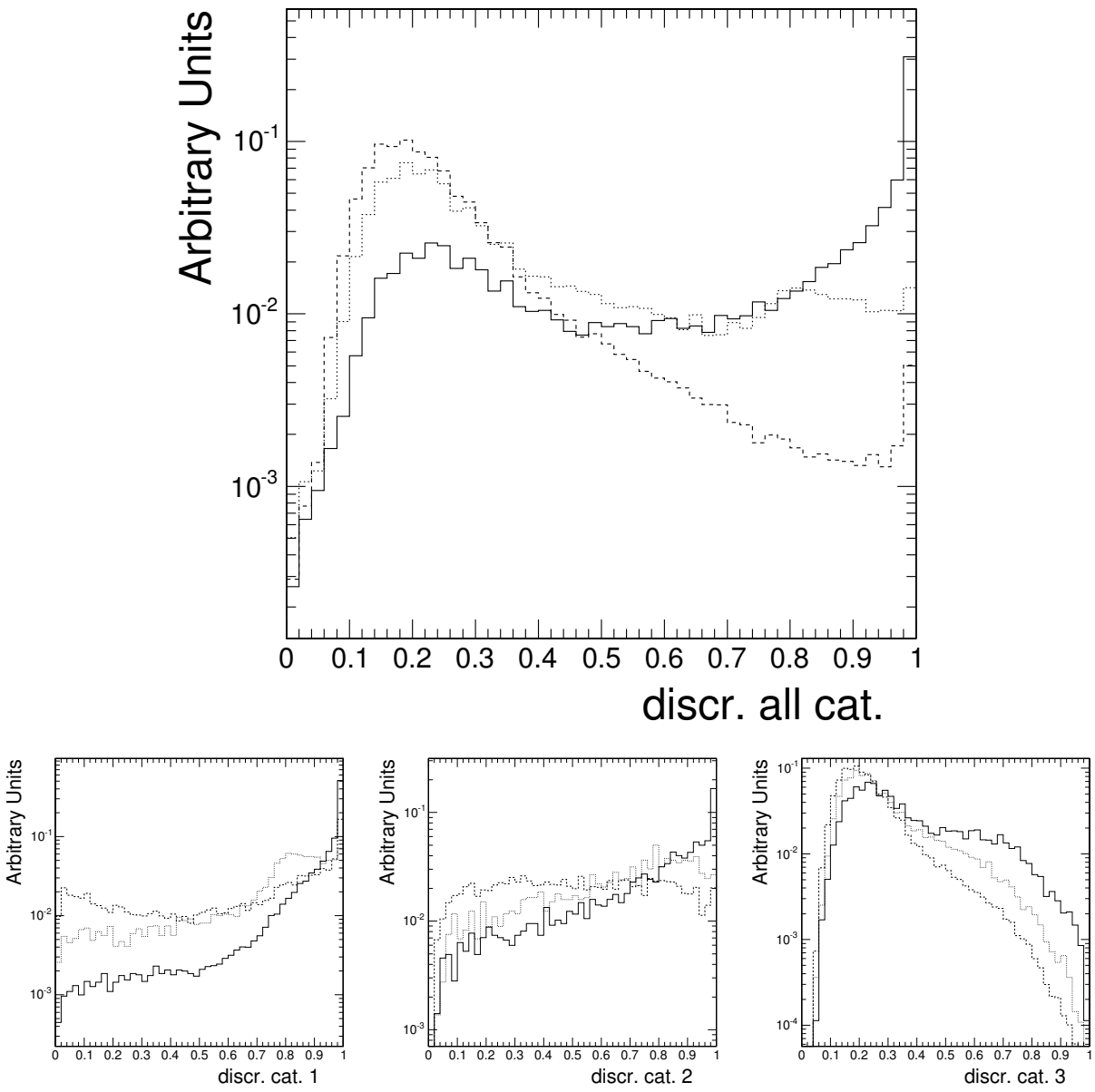


Figure 5: The distribution of the discriminator for all vertex categories (top), and separately for category "RecoVertex" (bottom left), "PseudoVertex" (bottom middle) and "NoVertex" (bottom right) for b-jets (solid), c-jets (dotted) and udsg-jets (dashed).

6 Performance of the Algorithm

In this section, the offline performance of the algorithm is shown. All event samples used for these studies were simulated with pile-up events as expected for a luminosity of $2 \cdot 10^{33} cm^{-2} s^{-1}$. The diagrams shown in the following have been obtained by scanning the cut on the discriminator. Fig. 6 shows the efficiencies to tag jets of different flavours as b-jet versus the cut on the combined discriminating variable.

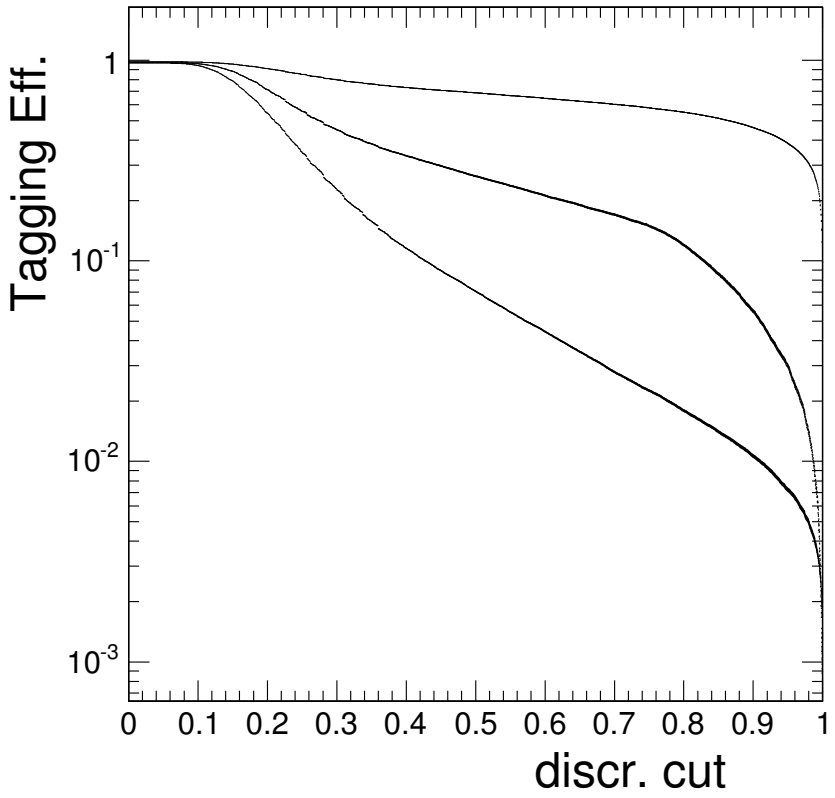


Figure 6: Efficiencies to tag a jet as b-jet versus the cut on the discriminator output for b-jets (top curve), c-jets (middle curve) and udsg-jets (bottom curve).

Fig. 7 shows the non-b jet mistag probability versus the b-jet tagging efficiency in the barrel ($|\eta| < 1.4$) and forward part of the detector ($1.4 < |\eta| < 2.4$) for QCD jets with transverse momenta between 50 GeV/c and 80 GeV/c.

Fig. 9 shows the performance as obtained from a sample of semi-leptonic $\mathrm{t\bar{t}}$ events for jets with transverse momenta larger than 30 GeV/c. This transverse momentum cut corresponds to a typical requirement in $\mathrm{t\bar{t}}$ analyses.

To compare the effect of the different definitions of the true jet flavour, the "physics" definition and the "algorithmic" definition as introduced in Section 1, Figures 8 and 10 show the same kind of performance plot for these event samples following the "algorithmic" definition. In the QCD sample, as expected, the main difference is for gluon jets, where it makes an enormous difference whether splitting to heavy quark pairs is included or not. In the "algorithmic" definition, jets originating from light quarks and gluons look very similar. Smaller effects come from the fact, that also the heavy flavour jets (b and c) have a somewhat different composition (because of the "contamination" from gluon or light quark jets with gluon splitting to heavy flavours inside) and the different momentum spectrum of the jets considered, because jets originating from (hard) final state radiation are counted separately.

As can be clearly seen in these figures, the b-tagging performance in the central part of the detector is significantly better than in the forward part. At a b-tagging efficiency of 60%, jets from light quarks can be suppressed by a factor of about 100 in the central part of the detector, whereas in the forward part the b-tagging efficiency is reduced to about 50% for the same light quark jet rejection. The suppression of gluon jets is more difficult, especially for tighter cuts, when the gluon splitting into heavy flavours ($\mathrm{c\bar{c}}$ and $\mathrm{b\bar{b}}$) becomes an almost irreducible background.

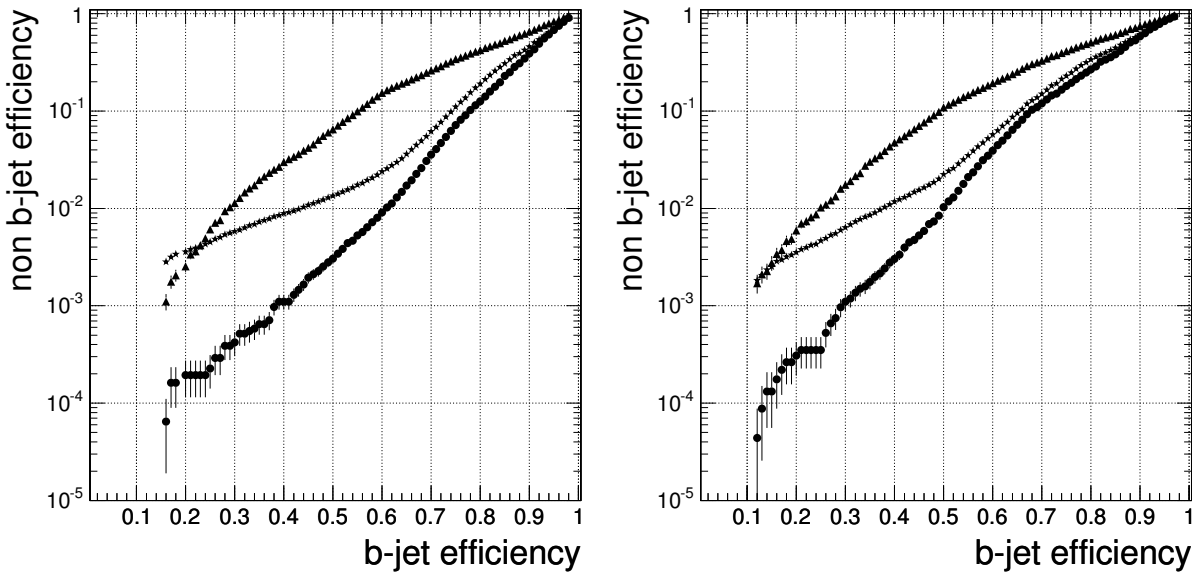


Figure 7: Non-b jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles), uds-jets (circles) and g-jets (stars) obtained for jets in a QCD sample for transverse jet momenta between 50 GeV/c and 80 GeV/c in the barrel ($|\eta| < 1.4$, left) and forward ($1.4 < |\eta| < 2.4$, right) regions of the detector.

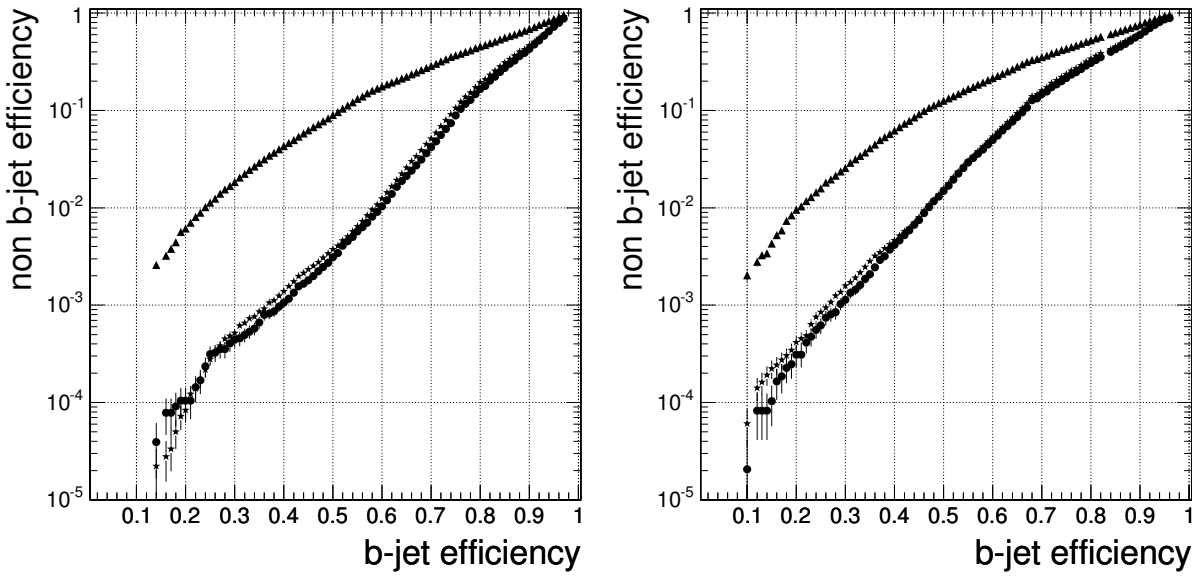


Figure 8: Non-b jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles), uds-jets (circles) and g-jets (stars) obtained for jets in a QCD sample for transverse jet momenta between 50 GeV/c and 80 GeV/c in the barrel ($|\eta| < 1.4$, left) and forward ($1.4 < |\eta| < 2.4$, right) regions of the detector following the "algorithmic" definition of the true jet flavour.

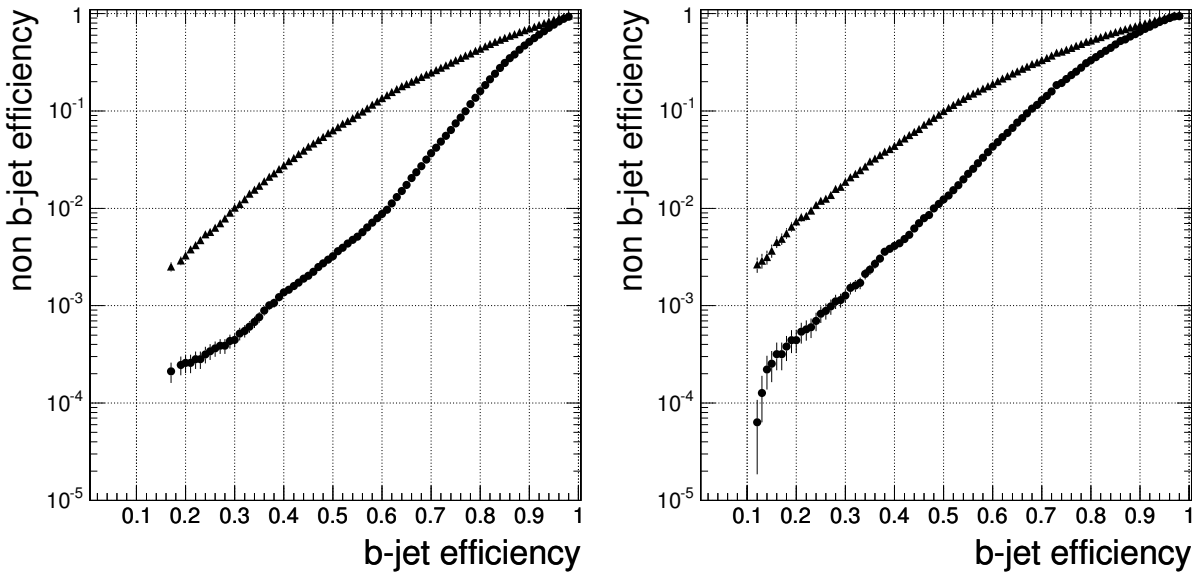


Figure 9: Non-b jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles) and uds-jets (circles) obtained for jets in a sample of semi-leptonic $\mathrm{t\bar{t}}$ events for jets with transverse momentum larger than 30 GeV/c in the barrel region ($|\eta| < 1.4$, left) and forward region ($1.4 < |\eta| < 2.4$, right) of the detector for the combined secondary vertex algorithm.

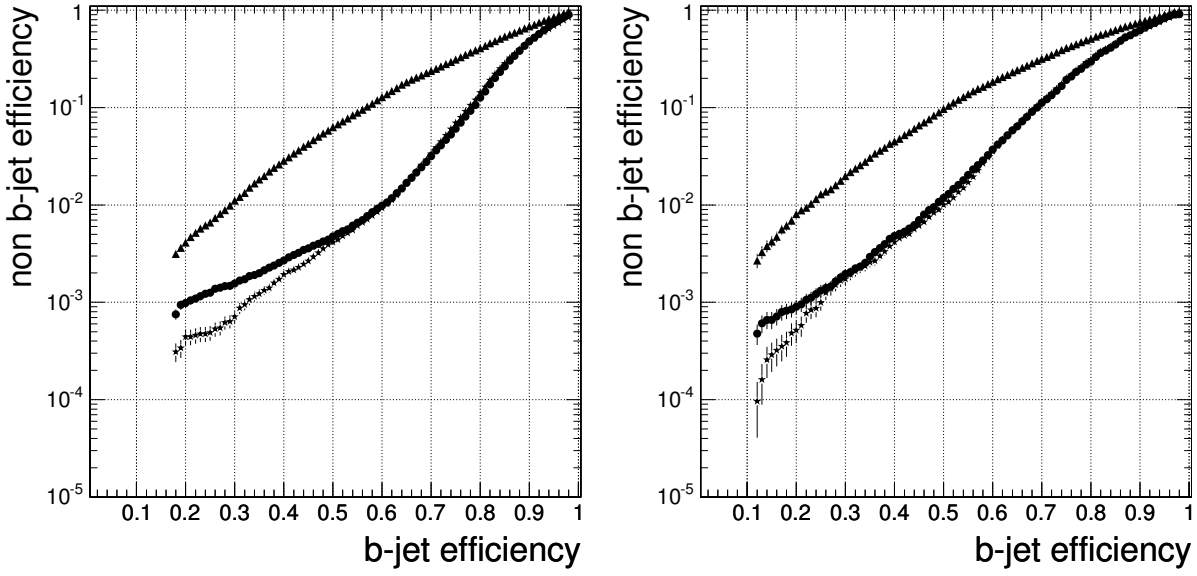


Figure 10: Non-b jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles), uds-jets (circles) and gluon jets (stars) obtained for jets in a sample of semi-leptonic $\mathrm{t\bar{t}}$ events for jets with transverse momentum larger than 30 GeV/c in the barrel region ($|\eta| < 1.4$, left) and forward region ($1.4 < |\eta| < 2.4$, right) of the detector for the combined secondary vertex algorithm following the "algorithmic" definition of the true jet flavour.

The dependence of the b-tagging performance on the jet transverse momentum and on the pseudorapidity is shown in Fig. 11. The non-b jet mistagging probability for a fixed b-tagging efficiency of 50% is shown versus the transverse momentum (left) and pseudorapidity (right) of the underlying initial parton.

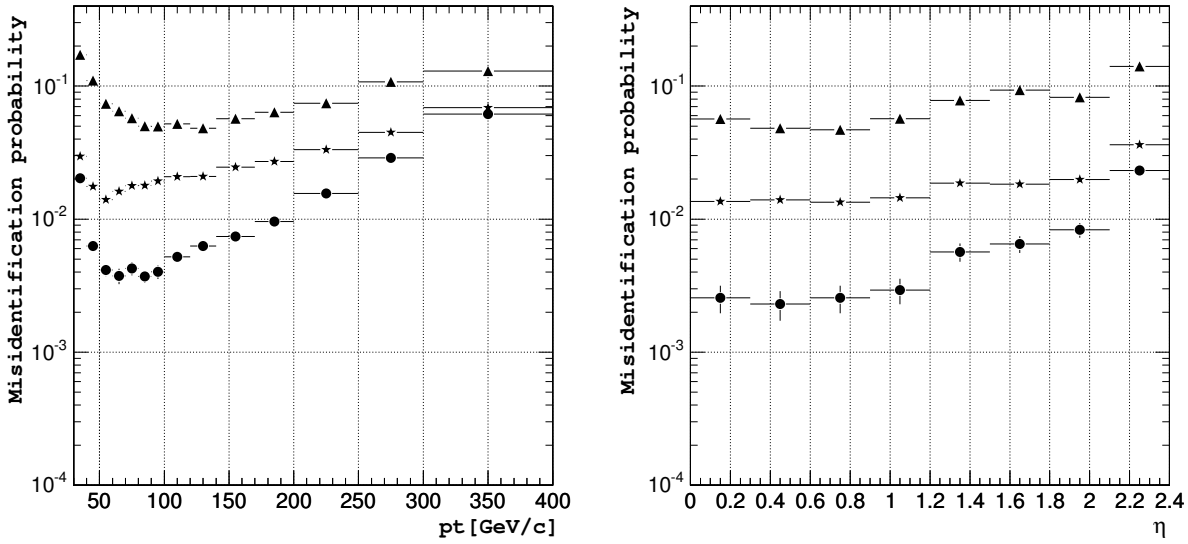


Figure 11: The non-b jet mistagging probabilities for a fixed b-jet tagging efficiency of 0.5 as a function of the jet transverse momentum (left, jets with $|\eta| < 2.4$ were considered) and pseudorapidity (right, jets with transverse momenta between 50 GeV/c and 80 GeV/c were considered) for c-jets (triangles), uds-jets (circles) and gluon jets (stars) obtained for jets in an event sample of QCD events for the combined secondary vertex algorithm.

Figures 12, 13, 14, and 15 show the non-b jet mistag probability versus the b-jet tagging efficiency in the barrel ($|\eta| < 1.4$) and forward part ($1.4 < |\eta| < 2.4$) of the detector for QCD jets in various ranges of transverse momentum, ranging from about 30 GeV/c to more than 200 GeV/c.

The main limiting factors to the light quark and gluon jets rejection come from V^0 decays, interactions in the detector material, gluon splitting into charm and b-quark pairs and limited detector resolution. Especially for gluon jets, the splitting into heavy flavour quark pairs is a serious limitation, particularly for working points with reduced b-tagging efficiency and large rejection factors against non-b jets. In this region most of the mistagged gluon jets are jets where the gluon has undergone splitting into b- or c-quarks. In the case of c-jets the rejection is obviously limited by the lifetime of c-hadrons giving signatures similar to those for b-hadron decays.

As can be seen in Fig. 11, the optimal performance is obtained in the central region of the detector and for transverse jet momenta of about 60-90 GeV/c. The performance degrades at larger pseudorapidities mainly because of the increased material budget traversed and degraded detector resolution. The slow degradation for larger transverse momenta is caused by an increased splitting rate of gluons to heavy quark pairs, an increased track multiplicity from fragmentation and more difficult pattern recognition in dense jets. The steep fall of performance towards lower transverse momenta is mainly due to increased multiple scattering, resulting in a worse separation of primary and secondary vertices, and also limitations in the jet reconstruction.

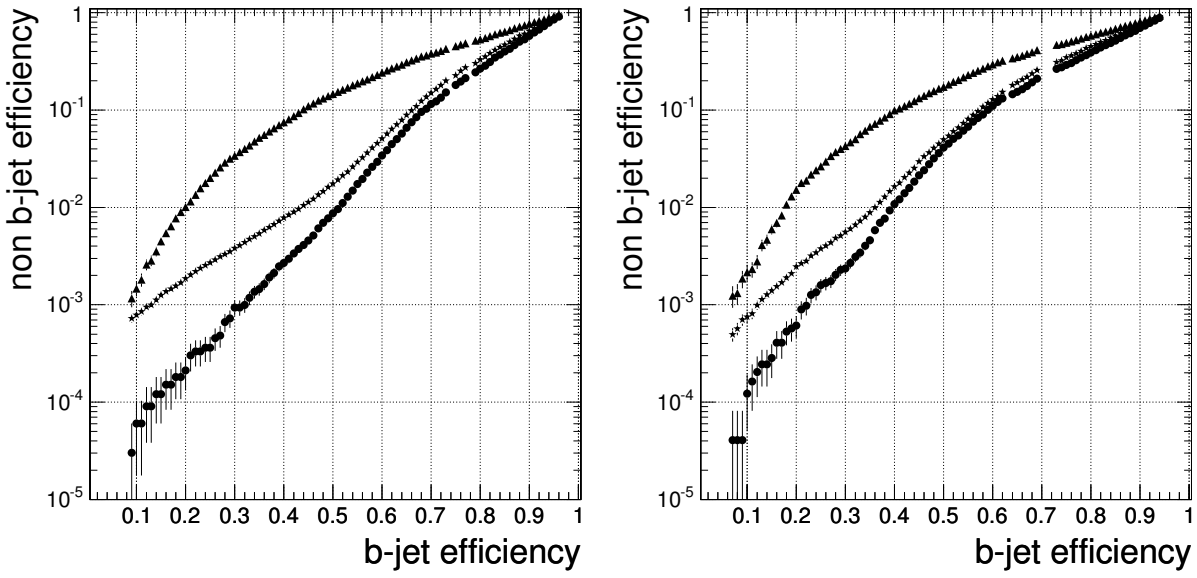


Figure 12: Non-b jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles), uds-jets (circles) and g-jets (stars) obtained for jets in a QCD sample for transverse jet momenta between 30 GeV/c and 50 GeV/c in the barrel ($|\eta| < 1.4$, left) and forward ($1.4 < |\eta| < 2.4$, right) regions of the detector.

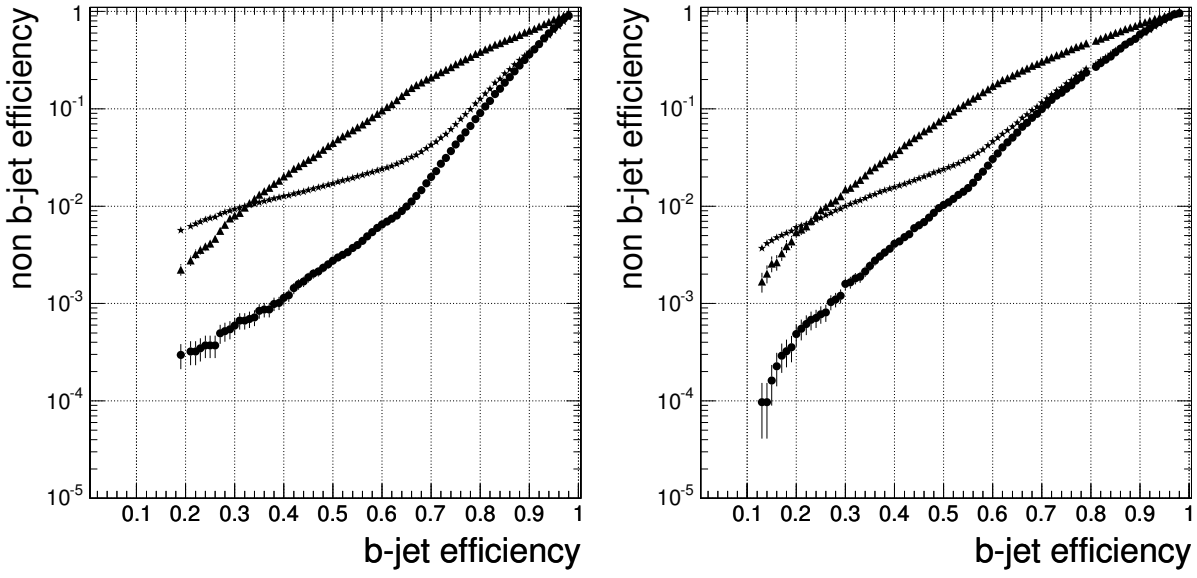


Figure 13: Non-b jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles), uds-jets (circles) and g-jets (stars) obtained for jets in a QCD sample for transverse jet momenta between 80 GeV/c and 120 GeV/c in the barrel ($|\eta| < 1.4$, left) and forward ($1.4 < |\eta| < 2.4$, right) regions of the detector.

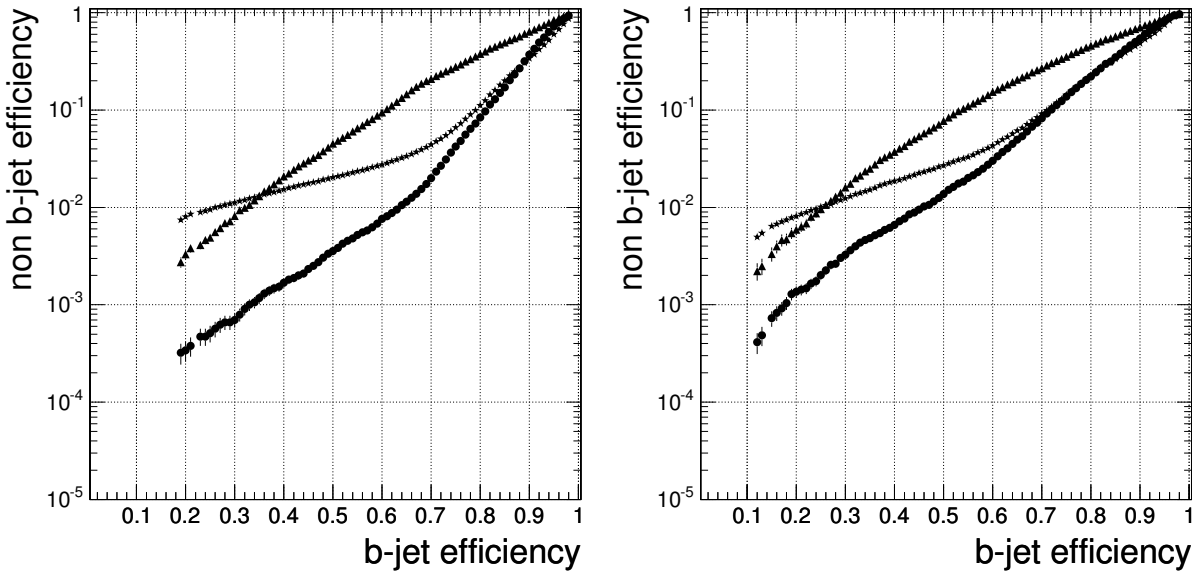


Figure 14: Non-b jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles), uds-jets (circles) and g-jets (stars) obtained for jets in a QCD sample for transverse jet momenta between 120 GeV/c and 170 GeV/c in the barrel ($|\eta| < 1.4$, left) and forward ($1.4 < |\eta| < 2.4$, right) regions of the detector.

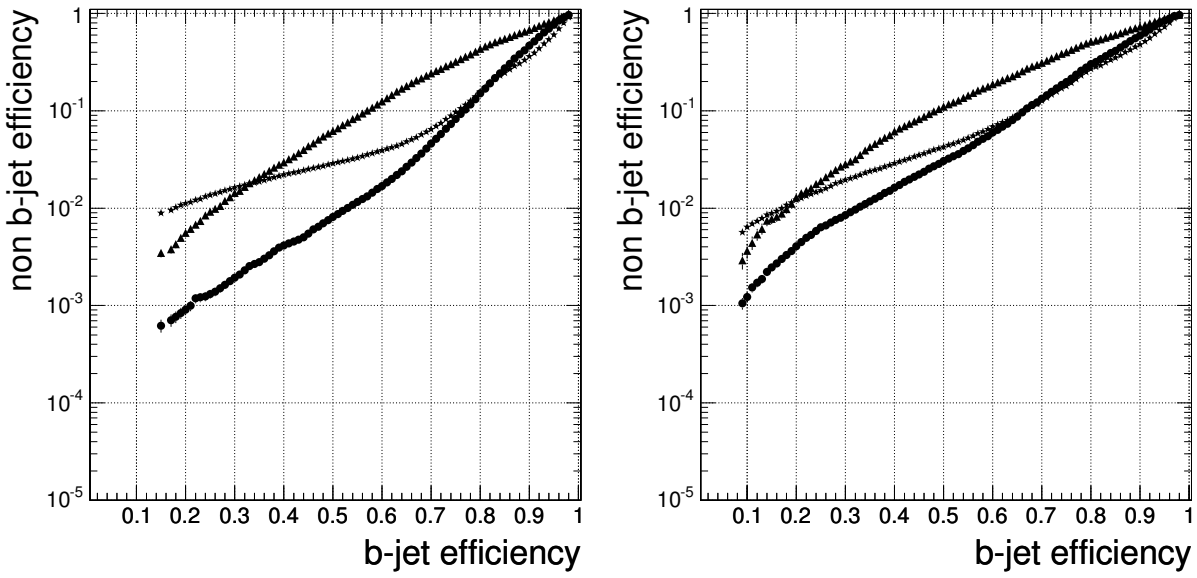


Figure 15: Non-b jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles), uds-jets (circles) and g-jets (stars) obtained for jets in a QCD sample for transverse jet momenta above 170 GeV/c in the barrel ($|\eta| < 1.4$, left) and forward ($1.4 < |\eta| < 2.4$, right) regions of the detector.

7 Robustness of the Algorithm

The performance studies shown in the previous section were carried out under the assumption of a fully equipped and perfectly aligned detector. This does not correspond, however, to the real situation one will face once the detector will be running. At the beginning of data taking some components of the detector might be missing and perfect alignment can never be reached. In this section the focus will be on studies of the performance expected for a not perfectly aligned tracking detector. Since the pixel and silicon strip detectors are extremely precise devices with spatial resolutions of about 10 μm for the inner layers, aligning the tracker modules with a precision at the same level as the spatial resolution is a particular challenge. Two misalignment scenarios have been defined, *"first data taking"* (accuracy as expected from mounting precision, the laser system and some track based alignment for the pixel detectors as expected after an integrated luminosity of less than 1 fb^{-1}) and *"long term"* (ultimate precision as expected from the track based alignment after a period of data taking with sufficient integrated luminosity, about 10 fb^{-1}). The assumed accuracy for the *long term* scenario is typically about a factor 10 better than for the *first data taking* scenario, apart for the pixel detectors where alignment precisions of about 10 μm are assumed for both scenarios. The alignment errors are added in quadrature to the nominal hit errors. Details of misalignment scenarios and procedures can be found in [5]; effects on track and vertex reconstruction are described in [6]. The combined b-tagging performance for a perfectly aligned detector and for the two misalignment scenarios described above is shown in Fig. 16. For these studies, the probability density functions (PDFs) of the variables entering into the final discriminating variable were the same as those used for the perfectly aligned detector. Adapting these PDFs might slightly improve the performance for the misalignment scenarios. Since the level of misalignment will be difficult to estimate during the first period of data taking, however, it seems more realistic to use the same PDFs for all cases.

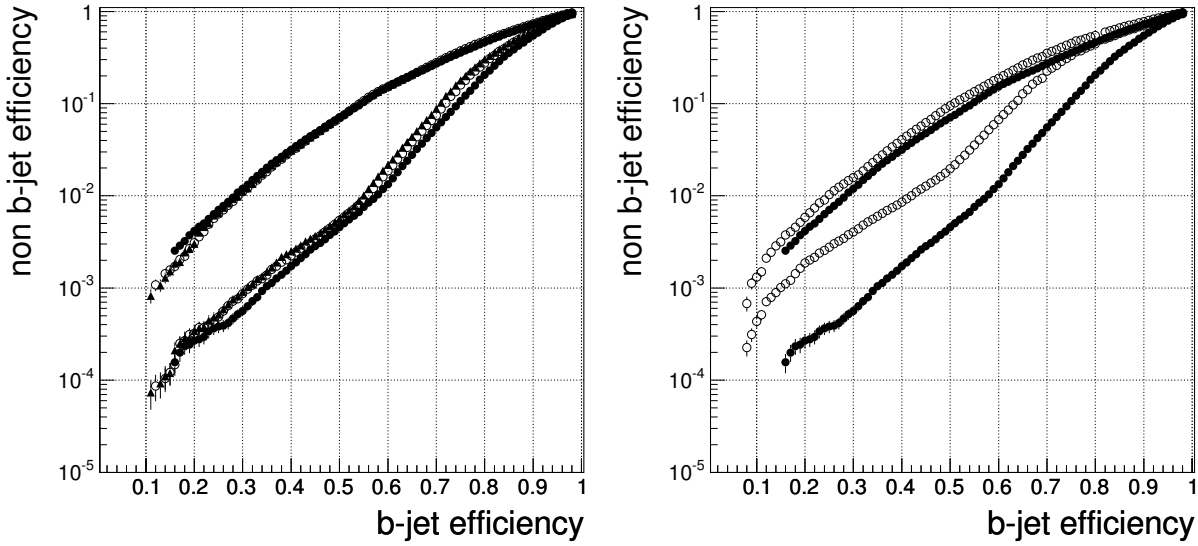


Figure 16: Non-b jet mistagging probability versus b-jet tagging efficiency for different misalignment scenarios. In the first plot (left) the standard scenarios considered are shown: perfect alignment (full circles), long term (open circles) and first data taking (triangles). The second plot (right) shows a comparison between a perfectly aligned detector (full circles) and the misalignment based on the first data taking scenario with additional scaling of the pixel detector misalignment by a factor of three (open circles). In both plots, the upper and lower set of curves correspond to charm-jets and uds-jets, respectively. The semi-leptonic $\mathrm{t\bar{t}}$ sample was used for these studies. Jets with transverse momenta greater than 30 GeV/c within the full tracker acceptance ($|\eta| < 2.4$) were considered.

The observed degradation of the b-tagging performance for light quark (u,d and s) and gluon jets is very similar for both misalignment scenarios. This can be explained by the fact, that both misalignment scenarios assume a similar alignment precision of the pixel detectors, affecting in the same way the track impact parameter resolution which is determined mainly by the first measurement points closest to the interaction region. Furthermore, for hadrons with typical momenta of some GeV/c as they are present in jets, the error on the track impact parameter due to multiple scattering is bigger than the alignment error assumed here. Therefore, the loss of performance with respect to the perfectly aligned detector is acceptable. It can be seen that there is essentially no effect on the

mistagging efficiency of charm jets if the tagging efficiency is kept constant. This can be explained by the fact, that for a large fraction of c-jets tagged as b-jets the real lifetime of charm hadrons is the origin of the mistag. Variables sensitive to the lifetimes of heavy hadrons thus vary in a very similar way for c- and b-jets when applying the misalignment.

To further illustrate the dominant role of the alignment of the pixel detectors, in Fig. 16 the performance is also shown comparing a perfectly aligned detector and a misalignment scenario based on the *"first data taking"* scenario, and movements, rotations etc. of the pixel detectors have been scaled by an additional factor of three, resulting in typical movements of the pixel detectors by about 30 μm. After this scaling, the degradation in b-tagging performance becomes clearly visible. Also in this case, light quark jets are much more affected than charm quark jets. However, for large misalignments of the pixel detectors, an adaption of the PDFs would be desirable because the distributions of the input variables change significantly.

There are also other effects that might have serious implications on b-tagging, e.g. a reduced track reconstruction efficiency or degraded track parameter resolutions because of improper modelling of the detector material or missing parts of the tracking detector (especially missing parts of the pixel detectors). These will be studied in the near future.

8 Summary and Outlook

In this Note, a b-tagging algorithm based on inclusive secondary vertex reconstruction in jets has been presented. Depending on the transverse momentum and pseudorapidity of jets, rejection factors against light quark jets up to 500 can be reached at a b-tagging efficiency of 50%. Its performance and robustness against detector misalignment has also been demonstrated.

References

[1] A. Heister et al., Measurements of Jets with the CMS Detector at the LHC, **CMS Note in preparation**

[2] W. Adam et al., Track Reconstruction in the CMS Tracker, **CMS Note in preparation**

[3] T. Speer et al., Vertex Fitting in the CMS Tracker, **CMS Note in preparation**

[4] C. Piasecki, C. Weiser et al., Inclusive Secondary Vertex Reconstruction in Jets, **CMS Note in preparation**

[5] T. Lampen et al., Simulation of Misalignment Scenarios for CMS Tracking Devices, **CMS Note 2006/008**

[6] L. Barbone et al., Impact of Silicon Tracker Misalignment on Track and Vertex Reconstruction, **CMS Note in preparation**