b-Jet Identification in the CMS Experiment

The CMS Collaboration

Abstract

The identification of jets associated with the production of $b$ quarks is an essential tool both for the measurement of standard-model processes and in the search for physics beyond this model at the Large Hadron Collider. The CMS experiment has developed a variety of algorithms that use the impact parameters of charged-particle tracks, the properties of reconstructed decay vertices, the presence of a lepton or combinations of these quantities to select samples of jets with different $b$ purities. Proton-proton collisions recorded in 2011 and corresponding to an integrated luminosity of 4.7 fb$^{-1}$ have been used to compare the quality of the reconstruction with expectations from simulation. The performance of the algorithms in terms of efficiency and misidentification probability has been measured from multijet events.
1 Introduction

Jets that arise from bottom-quark hadronisation and decay (b jets) are present in a wide range of physics processes of interest, such as the decay of top quarks, Higgs bosons, and various supersymmetric particles. The ability to accurately identify b jets is vital in reducing the otherwise overwhelming background to these channels from processes involving jets from gluons (g) and light-flavour quarks (u, d, s) and from c-quark fragmentation.

Properties of the bottom and, to a lesser extent, the charm quark can be used to identify the hadronic jets into which they fragment. These properties include their hard fragmentation functions and the relatively large mass and long lifetime of the heavy flavour hadrons. Their semileptonic decays can be exploited as well. The Compact Muon Solenoid (CMS) detector, with its precise charged particle tracking system and robust lepton identification is well matched to the task of b-jet identification (b tagging). First physics results using b tagging have already been obtained from the first LHC data samples [1–3] and many more are expected.

In the first part of this paper (Sec. 3) several algorithms for b tagging that take advantage of the properties of b jets are presented. The distributions of the relevant observables are compared between p-p collision data collected in 2011 at a center-of-mass energy of 7 TeV and simulation. The robustness of the algorithms with respect to running conditions such as the status of the alignment and the presence of multiple p-p collisions (pile-up) is also shown.

Any physics analysis using b-jet identification needs information about the efficiency and misidentification rate of the chosen algorithm and operating point [4]. In general, these values are a function of the transverse momentum and pseudorapidity of a jet. The performance of the algorithms can also depend on parameters such as the efficiency of the track reconstruction, the resolution of the reconstructed track parameters or the track density in a jet. While the CMS physics and detector simulation reproduces the performance of the detector to a remarkable degree of precision, it is difficult to control all the parameters relevant for b tagging. Therefore it is essential to measure the performance of the algorithms directly from data.

In the second part of this paper (Sec. 4), several methods are presented to measure the b-tagging efficiency from multijet samples. These include using the kinematic properties of muons associated to jets (Sec. 4.1), applying multiple taggers with weakly correlated efficiencies (Sec. 4.2), and using results of a specific tagging algorithm as reference (Sec. 4.3). These results are then summarized and are compared with the efficiencies obtained in the analysis of t̅t events [5]. The misidentification efficiency of light flavour jets (mistagging rate) is also determined from data. This measurement is presented in Sec. 5.

2 Detector and data samples

The central feature of the CMS apparatus is a superconducting solenoid, of 6 m internal diameter, which provides a magnetic field of 3.8 T. Within the field volume are the silicon tracker, the crystal electromagnetic calorimeter and the brass/scintillator hadron calorimeter. Muons are measured in gas-ionization detectors embedded in the steel return yoke.

CMS uses a right-handed coordinate system, with the origin at the nominal interaction point, the x-axis pointing to the centre of the LHC and the z-axis along the counterclockwise-beam direction. The polar angle, θ, is measured from the positive z-axis and the azimuthal angle, φ, is measured in the x-y plane. The pseudorapidity, η, is defined as − ln tan θ/2. A more detailed description of CMS can be found elsewhere [6].
The most relevant detector elements for the identification of b jets and the measurement of algorithm performance are the tracking system and the muon detectors. The inner tracker measures charged particles within the pseudorapidity range $|\eta| < 2.5$ and consists of 1440 silicon pixel and 15 148 silicon strip detector modules. The pixel modules are arranged in three cylindrical layers in the central part of CMS and two endcap disks at each side of the interaction point. The silicon strip detector comprises two cylindrical barrel detectors with a total of 10 layers and two endcap systems with a total of 12 layers at positive and negative $z$. The tracking system provides an impact parameter (IP) resolution of $\sim 15 \mu m$ and a transverse momentum ($p_T$) resolution of about 1.5% for 100 GeV particles.

Muons are measured in the pseudorapidity range $|\eta| < 2.4$, with detection planes made using three technologies: drift tubes, cathode strip chambers, and resistive plate chambers. A global muon reconstruction is achieved by matching segments found in the muon chambers to the tracks measured in the silicon tracker. This procedure yields muon candidates of high purity with a transverse momentum resolution between 1 and 5%, for $p_T$ values up to 1 TeV.

The data samples used for this article have been selected using a two-tier trigger system comprised of a first level (L1) based on custom hardware processors and a High Level Trigger (HLT) farm. They correspond to an integrated luminosity of 4.7 fb$^{-1}$ unless mentioned otherwise. Samples of inclusive multijet events were collected using jet triggers with $p_T$ thresholds of 30 to 300 GeV. For efficiency measurements, dedicated triggers were used which required the presence of at least two jets with $p_T$ thresholds ranging from 20 to 110 GeV. One of these jets had to include a muon with $p_T > 5$ GeV within a cone of $\Delta R = 0.4$ around the jet axis, where $\Delta R$ is defined as $\sqrt{\Delta \phi^2 + \Delta \eta^2}$.

Simulated samples of multijet events were generated with PYTHIA 6.424 [7] using the Z2 tune [8]. For b-tagging efficiency studies, dedicated multijet samples have been produced with the explicit requirement of a muon in the final state. The pair production of top quarks has been simulated with the MADGRAPH 5.1.3 event generator [9], interfaced with PYTHIA for the showering of the partons. The generated events were passed through the full simulation of the CMS detector based on GEANT4 [10]. In the simulation, a reconstructed jet is matched with a generated parton if the direction between the parton and the jet is within a cone of radius $\Delta R = 0.3$. The jet is then assigned the flavour of the parton. Should more than one parton be matched to a given jet, the flavour assigned is that of the heaviest parton, i.e. the b flavour is given priority over the c flavor, which in turn is given priority over light flavours. According to this definition jets originating from gluon splitting to b$b$, which constitute an irreducible background for all tagging algorithms, are classified as b jets.

### 3 Algorithms for b-jet identification

A variety of reconstructed objects – tracks, vertices and identified leptons – can be used to build observables which discriminate between b and light-flavour jets. Several simple and robust algorithms use just a single observable while others combine several of these measures to achieve a higher discrimination power. Each of these algorithms yields a single discriminator value for each jet. Minimum thresholds on these discriminators define loose (“L”), medium (“M”) and tight (“T”) working points with a nominal misidentification probability for light-flavour jets of 10%, 1% and 0.1%, respectively, at an average jet $p_T$ of about 80 GeV. Throughout this paper tagging criteria will be labeled with the letter characterizing the operating point appended to the acronym of one of the algorithms described in Sec. 3.2 and 3.3.

After a short description of the reconstructed objects used as inputs, details on the tagging al-
3.1 Reconstructed objects

The results reported in this paper apply to jets which have been reconstructed using the particle-flow algorithm [11, 12]. This algorithm combines information from all subdetectors to create a consistent list of reconstructed particles for an entire event. The particles are then clustered into jets using the anti-$k_T$ clustering algorithm [13] with a size parameter of 0.5. The raw jet energies were corrected to obtain a uniform response in $\eta$ and an absolute calibration in $p_T$. Details on the performance of the jet reconstruction can be found in [14]. While particle-flow jets are used as the default, the b-tagging algorithms can be applied in a straightforward way to any reconstructed jet.

Each of the algorithms described in the next section uses the measured kinematic properties of charged particles in a jet, some of which may also be identified as leptons. The trajectories of these particles are reconstructed in the CMS tracking system in an iterative procedure using a standard Kalman filter based method. Details on the pattern recognition, the track-parameter estimation and the tracking performance in proton-proton collisions can be found in Ref. [15].

Primary vertex candidates are selected by clustering reconstructed tracks based on the $z$ coordinate of their closest approach to the beam line. An adaptive vertex fit [16] is then used to estimate the vertex position using a sample of tracks compatible with originating from the interaction region. Among the primary vertices found in this way, the one with the highest $\sum_{\text{Tracks}} p_T^2$ is selected as a candidate for the origin of the hard interaction.

The b-tagging algorithms require a sample of well-reconstructed tracks of high purity, and specific requirements are imposed in addition to the selection applied in the tracking step. The fraction of fake or poorly reconstructed tracks is reduced by requiring a transverse momentum of at least 1 GeV, at least eight hits to be associated to the track and a good fit quality, $\chi^2/\text{ndof} < 5$, where ndof are the number of degrees of freedom in the track fit. Since track measurements in the vicinity of the interaction vertex contain most of the discriminating power at least two hits are required in the pixel system. A loose selection on the track impact parameters is used to further increase the purity and to reduce the contamination by decay products of long-lived particles like $K_0^*$'s. The transverse and longitudinal impact parameters, $d_{xy}$ and $d_z$, are defined as the transverse and longitudinal distance to the primary vertex at the point of closest approach to the beam line. Their norms have to be smaller than 0.2 cm and 17 cm, respectively. Tracks are associated to jets in a cone $\Delta R < 0.5$ around the jet axis. At the point of closest approach the distance to the jet axis has to be smaller than 700 $\mu$m and this point has to be within 5 cm of the primary vertex. This sample of associated tracks is the basis for all algorithms which use impact parameters for discrimination. Properties of the tracks and the average multiplicity after the selection (except for the variable plotted) are shown in Fig. 1. The distributions show good agreement with the expectations from simulation. The track multiplicity and the lower part of momentum spectrum are particularly sensitive to the modeling of the particle multiplicity and kinematics by the Monte Carlo generator. The corresponding differences between data and simulation that are not due to detector performance and indirect effects are also seen in distributions like the number of hits in the innermost layers. In Fig. 1 and the following figures simulated events with gluon splitting to $b\bar{b}$ are shown as a special category: the $b$ jets in these events tend to be close in space and can be merged by the clustering algorithm, resulting in a higher average track multiplicity per jet.
Figure 1: Track properties after basic selection (except for the variable plotted): (a) number of hits in the pixel system, (b) transverse momentum, (c) distance to the jet axis. The average number of tracks passing the basic selection is shown in (d) as a function of the transverse momentum of the jet. In figures (a)-(c) the distributions from simulation have been renormalized to match the counts in data. The points correspond to data while the stacked, coloured histograms indicate the contributions of different components as predicted by the simulation of multijet ("QCD") samples. Simulated events with gluon splitting to b quarks ("b from gluon splitting") are distinguished from other b production processes ("b quark"). In each histogram, the rightmost bins include all events above the upper boundary. The sample corresponds to a trigger selection at jet $p_T > 60$ GeV.

In order to minimize the combinatorial complexity in the reconstruction of secondary vertices, which is more challenging in the presence of multiple proton-proton interactions, a more restrictive track selection is applied. Tracks have to be within a cone of $\Delta R = 0.3$ around the jet axis with a maximal distance to this axis of 0.2 cm and pass a "high purity" criterion of the track reconstruction algorithm [15], which uses the normalized $\chi^2$ of the track fit, the track length and impact parameter information to optimize the purity for each of the iterations in track reconstruction. The vertex finding starts with the tracks defined by this selection and proceeds in an iterative way. A vertex candidate is identified by applying an adaptive vertex fit [16], which is robust in the presence of outliers. The fit estimates the vertex position and assigns a weight between 0 and 1 to each track based on its compatibility with the vertex. All tracks with weights $> 0.5$ are then removed from the sample and the fit procedure is repeated, until no new vertex candidate can be found. In the first iteration the interaction region is used as a constraint in order to identify the prompt tracks in the jet; the following iterations produce
3.2 Identification using track impact parameters

The impact parameter (IP) of a track with respect to the primary vertex can be used to distinguish decay products of a b hadron from prompt tracks. The IP is calculated in three dimensions, taking advantage of the excellent resolution of the pixel detector along the z axis. The impact parameters are signed according to the scalar product of the vector pointing from the primary vertex to the point of closest approach with the jet direction: tracks originating from the decay of particles traveling along the jet axis will tend to have positive IP values while the impact parameters of prompt tracks can have positive or negative signs. The resolution on the impact parameter depends strongly on $p_T$ and $\eta$ of a track. Therefore the impact parameter significance $S_{IP}$, defined as the ratio between the IP and its estimated uncertainty, is used as observable. The distribution of IP values and significances for all selected tracks are shown in Fig. 2. In general good agreement is observed with the exception of a small difference in the width of the core of the IP significance distribution.

The impact parameter significance alone has discriminating power between the decay products of b and non-b jets. The Track Counting (TC) algorithm sorts tracks in a jet by decreasing values of the IP significance. While the ranking tends to bias the values for the first track to high positive IP significances, the probability to have several tracks with high positive values is low for light-flavour jets. Therefore the two versions of the algorithm use the IP significance of the second and third ranked track as the discriminator value. In the following these two versions are called Track Counting High Efficiency (TCHE) and Track Counting High Purity (TCHP) algorithm, respectively. The discriminators for these two algorithms are shown in Fig. 3.

A natural extension of the TC algorithms is the combination of the IP information of several tracks in a jet. Two discriminators are computed: the jet Probability (JP) algorithm uses an estimate of the likelihood that all tracks associated to the jet come from the primary vertex while the jet B Probability (JBP) algorithm gives more weight to the tracks with the highest IP significance up to a maximum of four such tracks, matching the average number of reconstructed charged particles from B-hadron decays. The estimate for the likelihood, $P_{jet}$, is defined as

$$P_{jet} = \Pi \cdot \sum_{i=0}^{N-1} \frac{(-\ln \Pi)^i}{i!} \quad \text{with} \quad \Pi = \prod_{i=1}^{N} \max(P_{i}, 0.005),$$

(1)
where \( N \) is the number of tracks under consideration and \( P_i \) is the compatibility of track \( i \) with the primary vertex [17, 18]. The \( P_i \) are based on the probability density functions for the IP significance of prompt tracks. These functions are extracted from data for different track quality classes, using the shape of the negative part of the \( S_{IP} \) distribution. The cut-off parameter for \( P_i \) at 0.5% limits the impact of single, poorly reconstructed tracks with extremely low compatibility on the global estimate. The discriminators for the jet probability algorithms have been chosen to be proportional to \(-\ln P_{jet}\). Their distributions in data and simulation are shown in Fig. 4.

3.3 Identification using secondary vertices

The presence of a secondary decay vertex and kinematic variables associated with this vertex can be used to discriminate between \( b \) and non-\( b \) jets. These variables include the flight distance and direction, based on the vector between primary and secondary vertex, and various properties of the system of associated secondary tracks such as the multiplicity, the mass or the energy. Secondary-vertex candidates have to pass a set of requirements to enhance the \( b \) purity:

- vertices which could be compatible with the primary vertex are rejected by requiring that < 65% of the tracks associated with the candidate are shared with the primary

Figure 3: Discriminator values for the TCHE (left) and TCHP (right) algorithms. Conditions and symbols are the same as in Fig. 1.

Figure 4: Discriminator values for the JP (left) and JBP (right) algorithm. Conditions and symbols are the same as in Fig. 1. The small discontinuities in the distributions are due to the limit of the single track probability at 0.5%.

3.3 Identification using secondary vertices

The presence of a secondary decay vertex and kinematic variables associated with this vertex can be used to discriminate between \( b \) and non-\( b \) jets. These variables include the flight distance and direction, based on the vector between primary and secondary vertex, and various properties of the system of associated secondary tracks such as the multiplicity, the mass or the energy. Secondary-vertex candidates have to pass a set of requirements to enhance the \( b \) purity:

- vertices which could be compatible with the primary vertex are rejected by requiring that < 65% of the tracks associated with the candidate are shared with the primary

3.3 Identification using secondary vertices

The presence of a secondary decay vertex and kinematic variables associated with this vertex can be used to discriminate between \( b \) and non-\( b \) jets. These variables include the flight distance and direction, based on the vector between primary and secondary vertex, and various properties of the system of associated secondary tracks such as the multiplicity, the mass or the energy. Secondary-vertex candidates have to pass a set of requirements to enhance the \( b \) purity:

- vertices which could be compatible with the primary vertex are rejected by requiring that < 65% of the tracks associated with the candidate are shared with the primary

3.3 Identification using secondary vertices

The presence of a secondary decay vertex and kinematic variables associated with this vertex can be used to discriminate between \( b \) and non-\( b \) jets. These variables include the flight distance and direction, based on the vector between primary and secondary vertex, and various properties of the system of associated secondary tracks such as the multiplicity, the mass or the energy. Secondary-vertex candidates have to pass a set of requirements to enhance the \( b \) purity:

- vertices which could be compatible with the primary vertex are rejected by requiring that < 65% of the tracks associated with the candidate are shared with the primary

3.3 Identification using secondary vertices

The presence of a secondary decay vertex and kinematic variables associated with this vertex can be used to discriminate between \( b \) and non-\( b \) jets. These variables include the flight distance and direction, based on the vector between primary and secondary vertex, and various properties of the system of associated secondary tracks such as the multiplicity, the mass or the energy. Secondary-vertex candidates have to pass a set of requirements to enhance the \( b \) purity:

- vertices which could be compatible with the primary vertex are rejected by requiring that < 65% of the tracks associated with the candidate are shared with the primary

3.3 Identification using secondary vertices

The presence of a secondary decay vertex and kinematic variables associated with this vertex can be used to discriminate between \( b \) and non-\( b \) jets. These variables include the flight distance and direction, based on the vector between primary and secondary vertex, and various properties of the system of associated secondary tracks such as the multiplicity, the mass or the energy. Secondary-vertex candidates have to pass a set of requirements to enhance the \( b \) purity:

- vertices which could be compatible with the primary vertex are rejected by requiring that < 65% of the tracks associated with the candidate are shared with the primary

3.3 Identification using secondary vertices

The presence of a secondary decay vertex and kinematic variables associated with this vertex can be used to discriminate between \( b \) and non-\( b \) jets. These variables include the flight distance and direction, based on the vector between primary and secondary vertex, and various properties of the system of associated secondary tracks such as the multiplicity, the mass or the energy. Secondary-vertex candidates have to pass a set of requirements to enhance the \( b \) purity:

- vertices which could be compatible with the primary vertex are rejected by requiring that < 65% of the tracks associated with the candidate are shared with the primary
3.3 Identification using secondary vertices

vertex and that the significance of the radial distance between the two vertices exceeds $3\sigma$;

- the contamination by interaction vertices and decays of long-lived mesons is reduced by rejecting candidates with a radial distance of $>2.5$ cm with respect to the primary vertex or with masses compatible with the mass of $K^0$ or exceeding 6.5 GeV;

- the flight direction of all candidates has to be within a cone of $\Delta R < 0.5$ around the jet direction.

The Simple Secondary Vertex (SSV) algorithms use the significance of the flight distance (the ratio of the flight distance to its estimated uncertainty) as discriminating variable. If several vertices are present the one with the highest significance is used. Similar to the Track Counting algorithms there exist two versions optimized for different purity: the High Efficiency (SSVHE) version uses vertices with at least two associated tracks while for the High Purity (SSVHP) version at least three tracks are required. In Fig. 5 the flight distance significance and the mass associated with the vertex are shown.

A more complex approach involves the use of secondary vertices, together with track-based lifetime information. By using these additional variables, the Combined Secondary Vertex (CSV) algorithm provides discrimination even when no secondary vertices are found, so the maximum possible b-tagging efficiency is not limited by the secondary vertex reconstruction efficiency. In many cases, tracks with an IP significance $>2$ can be combined in a so-called “pseudo vertex”, allowing for the computation of a subset of secondary vertex based quantities even without an actual vertex fit. When even this is not possible, a “no vertex” category reverts simply to track-based variables that are combined in a way similar to that of the Jet Probability algorithm.

The following set of variables with high discriminating power and low correlations is used:

- the vertex category (real, “pseudo,” or “no vertex”);
- the 2D flight distance significance;
- the vertex mass;
- the number of tracks at the vertex;
- the ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet;
- the pseudo-rapidity of the tracks at the vertex with respect to the jet axis;

Figure 5: Properties of reconstructed decay vertices: significance of the flight distance (left) and mass (right). Conditions and symbols are the same as in Fig. 1.
• the 2D IP significance of the first track that raises the invariant mass above the charm threshold of 1.5 GeV when subsequently summing up tracks ordered by decreasing IP significance;
• the number of tracks in the jet;
• the 3D signed IP significances for each track in the jet.

Two likelihood ratios are built from these variables, used to discriminate between b and c jets and between b and light jets, and then combined with prior weights of 0.25 and 0.75, respectively. The distributions of the vertex multiplicity and of the CSV discriminator are presented in Fig. 6.

The performance of the algorithms described above is summarized in Fig. 7 in form of the non-b vs. b efficiency curves predicted by simulation. At a misidentification probability for light-flavour jets of 10% a b-jet tagging efficiency of $\approx 80 - 85\%$ is achieved. For the tight operating points at a mistagging rate of 0.1% typical efficiency values are $\approx 45\%$. Of all the algorithms, the CSV algorithm has the highest b efficiency when the light-flavour efficiency is less than 3%, above which the JPB algorithm has the highest efficiency. The CSV and the jet probability algorithms provide better c-jet rejection in the high purity region.

### 3.4 Impact of running conditions on b-jet identification

Even the simplest tagging algorithms depend on high tracking efficiency and a reliable estimation of track parameters and their uncertainties. Thus they are potentially sensitive to changes in the running conditions of the experiment. The robustness of the algorithms with respect to the most important of these effects have been studied: the residual uncertainty on the alignment of the tracking system and increased track densities due to multiple simultaneous pp interactions.

The alignment of the CMS tracker is performed using a mixture of tracks from cosmic rays and minimum bias collisions [19, 20]. In addition, the geometry of the tracker is monitored on a daily basis. During the 2011 data taking, the most important movements were observed between the two halves of the pixel barrel detector with discrete changes in the relative $z$ position of up to $30\mu m$. The sensitivity to misalignment was studied on simulated tt samples. With the current estimated accuracy of the positions of the sensitive elements no significant deterioration is observed with respect to a perfectly aligned detector. The effects of displacements between the two parts of the pixel barrel detector was studied by introducing artificial separa-
Figure 7: Performance curves obtained from simulation for the algorithms described in the text: light flavour (left) and c (right) efficiencies as a function of the b efficiency.

Due to the high instantaneous luminosity delivered by the LHC in 2011, the number of proton collisions taking place simultaneously in one bunch crossing was of the order of 5 to 20 depending on the time period. Although tracks from additional pile-up collisions strongly increase the track multiplicity in the event, the track selection is able to reject tracks from nearby primary vertices. The multiplicity distribution of selected tracks is almost independent from the number of primary vertices as shown in Fig. 9 left. The rejection of the additional tracks is mainly due to the cut on the distance of the tracks with respect to the jet axis, a selection criterion that is very efficient for the rejection of tracks from pile-up. The reconstruction of track parameters is also hardly affected; the distribution IP significance of the 2nd-highest track is stable (Fig. 9 right). Fig. 9 (left) indicates a slightly lower tracking efficiency in events with high pile-up. The impact of high pile-up on the b-tagging performance is illustrated in Fig. 10, which shows the light-flavour mistagging rate versus the b-tagging efficiency for the TCHP and SSVHP algorithms. The changes are concentrated in the regions of very high purity. In order to focus on the changes due to the b-tagging algorithms the performance curves have been compared using a jet $p_T$ threshold of 60 GeV at the generator level.

4 Efficiency measurements

For the b-tagging algorithms to be used in physics analyses, it is crucial to know the efficiency for each algorithm to select real b jets. There are a number of techniques that can be applied to CMS data to measure the efficiencies in situ, and thus reduce the reliance on simulations. However, the most important quantity is not necessarily the absolute performance of the identification algorithms, but how well the detector simulation models that performance. All of the efficiency measurements that are done with collision data are also replicated in simulated samples using Monte Carlo truth-level information to identify jet flavour, and the “scale factor,”...
Figure 8: Mistagging probabilities for non-b jets versus the efficiency for b jets for the TCHP (left) and SSVHP (right) b-tagging algorithms in $t\bar{t}$ production in simulated scenarios with an artificial separation of the two BPIX half barrels of 40, 80, 120 and 160 $\mu$m.

Figure 9: Changes in observables for different levels of pile-up in data. Left: number of tracks associated to a jet passing the selection cuts. Right: IP significance of the second-most significant track. The distributions are shown three ranges of primary vertex multiplicity.
4.1 Efficiency measurement with kinematic properties of muon jets

Some efficiency measurements are performed using samples that include a jet that has a muon within $\Delta R = 0.4$ from the jet axis (a “muon jet”). Because the semileptonic branching fraction of $b$ hadrons is significantly larger than that for other hadrons (about 11%, or 20% when $b \rightarrow c \rightarrow l$ cascade decays are included), these jets are more likely to arise from $b$ quarks than from another flavour. Muons can be identified very efficiently in the CMS detector, making it straightforward to collect samples of muon jets. These can be used to measure the performance of the lifetime-based tagging algorithms, as the efficiencies of the two $b$-jet identification techniques are largely uncorrelated. Secs. 4.1 and 4.2 describe efficiency measurements that use muon jets, while the technique of Sec. 4.3 makes use of a more generic dijet sample.

4.1 Efficiency measurement with kinematic properties of muon jets

Due to the large $b$-quark mass, the component of the momentum of the muon transverse to the jet axis, $p_{T\text{rel}}$, is larger for muons from $b$-hadron decays than for muons in light-flavour jets or from charm hadrons. In addition, the impact parameter of the muon track, calculated in three dimensions, is also larger for $b$ hadrons than for other hadrons. Both of these variables can thus be used as a discriminant in the $b$-tagging efficiency determination. In both cases, the discriminating power of the variable varies with the muon-jet $p_T$; $p_{T\text{rel}}$ does not discriminate well between flavours above a jet $p_T$ of about 160 GeV, while the IP distributions have less discrimination for jets with $p_T$ smaller than $\approx 100$ GeV. The “PtRel” and “IP3D” methods [21] rely on fits to the $p_{T\text{rel}}$ and IP distributions in the data with simulated spectra for the $b$ signal and charm+light background.

In the two methods, the $p_{T\text{rel}}$ and IP spectra for muon jets are modeled with simulated distributions that represent the spectra expected for different jet flavours to obtain the $b$-jet content of the sample. The efficiency for a particular tagger is obtained by measuring the fraction of muon jets that satisfy the requirements of the tagger. To make the treatment of the statistical uncertainty more straightforward, the muon-jet sample is separated into those that satisfy and those that fail the requirements of the tagger; these jets are referred to as “tagged” and “untagged.”

Figure 10: Light-flavour mistagging rate versus $b$-tagging efficiency for different pile-up scenarios. Left: TCHP algorithm. Right: SSVHP algorithm.

$S_{FB}$, the ratio of efficiencies measured in data and simulated samples, is calculated.
A dijet sample with high b-jet purity is obtained by requiring that events have exactly two reconstructed jets: the muon jet as defined above and another jet fulfilling the TCHPM b-tagging criterion (the “medium” operating point for the TCHP algorithm). Simulated Monte Carlo events are used to establish \( p_{T_{rel}} \) and IP spectra for muon jets resulting from the fragmentation of b, c, and light-flavour partons. Muons in light-flavour jets mostly arise from the decay of charged pions or kaons and from fake muons or hadronic punch-through in the calorimeters, effects that might not be modeled well in the simulation. The spectra for light-flavour jets from simulation can be validated against control samples of collision data. In Fig. 11 the distributions of \( p_{T_{rel}} \) (left) and IP (right) derived from the simulation are compared to the ones obtained in inclusive jet data by applying the same kinematic selection and track reconstruction quality as for the muon candidates. In order to measure the ability of the simulation to model the investigated spectra, we apply the same procedure to a sample of simulated inclusive jet events; the ratio of the shape of the distributions obtained in data and simulation is used as a correction factor for the spectra derived for soft muons from light jets in simulation.

![Figure 11: Comparison of distributions of muon \( p_{T_{rel}} \) for jets with \( p_T \) between 80 and 120 GeV (left) and of IP for jets with \( p_T \) between 160 and 320 GeV (right) as obtained from data and from simulated light flavour and charm jets scaled to the data.](image)

The fractions of each jet flavour in the dijet sample are extracted with a binned maximum likelihood fit using \( p_{T_{rel}} \) and IP templates for b, c and udsg jets derived from simulation or inclusive jet data. The fits are performed independently in the tagged and untagged subsamples of the muon jets. Results of representative fits are shown in Fig. 12 and Fig. 13.

From each fit the fractions of b jets (\( f_{b}^{\text{tag}} \), \( f_{b}^{\text{untag}} \)) are extracted from the data. With these fractions and the total yields of tagged and untagged muon jets (\( N_{\text{tag}}^{\text{data}} \), \( N_{\text{untag}}^{\text{data}} \)) the number of b jets in these samples are calculated, and the efficiency \( \epsilon_{b}^{\text{tag}} \) for tagging b jets in the data is inferred:

\[
\epsilon_{b}^{\text{tag}} = f_{b}^{\text{tag}} \cdot \frac{N_{\text{tag}}^{\text{data}}}{f_{b}^{\text{tag}} \cdot N_{\text{tag}}^{\text{data}} + f_{b}^{\text{untag}} \cdot N_{\text{untag}}^{\text{data}}}.
\]

To obtain \( SF_{b} \), the efficiency for tagging b jets in the simulation is obtained from jets that have been identified as b jets with Monte Carlo truth information.

### 4.2 Efficiency measurement with the System8 method

The “System8” method [22, 23] relies on the use of three tagging criteria with only weakly correlated efficiencies applied to multiple jets in event that contains a muon jet, and then the solution of a system of equations that relates the efficiencies for these criteria. A muon jet can be
4.2 Efficiency measurement with the System8 method

Figure 12: Fits of the muon $p_{Trel}$ distributions to b and non-b templates for muon jets that (a, b) pass or (c, d) fail the b-tagging criteria for JPL and CSVM. The muon-jet $p_T$ is between 80 and 120 GeV.

tagged as a b jet using either a lifetime tagger, or by requiring that the muon have large $p_{Trel}$. In this analysis, the requirement is $p_{Trel} > 0.8$ GeV. These two tagging criteria have efficiencies $\varepsilon_b^{tag}$ and $\varepsilon_b^{p_{Trel}}$, respectively, for b jets. Then, another jet in the event, denoted the “away-tag” jet, can be required to satisfy a lifetime-based criterion also. This requirement enriches the b content of the events, and thus makes it more likely that the muon jet is a b jet. Correlations between the efficiencies of the two tagging criteria are estimated from simulation. As $p_{Trel}$ provides less discrimination between jet flavours at higher jet energies, the System8 method loses sensitivity above jet $p_T > 120$ GeV.

With these criteria, eight quantities can be measured: the total number of muon jets in the sample $n$, the number of muon jets that pass the lifetime-tagging criterion $n^{Tag}$, the number of muon jets that pass the $p_{Trel}$ requirement $n^{p_{Trel}}$, the number of muon jets that pass both criteria $n^{Tag,p_{Trel}}$, and the same quantities for the subset of events in which the away jet is also tagged with the TCHPL criterion ($p$, $p^{Tag}$, $p^{p_{Trel}}$, $p^{Tag,p_{Trel}}$). The full sample and the away-tag sample are each composed of an unknown mix of b and non-b (labeled $c\ell$) jets ($n_b$, $n_{c\ell}$, $p_b$, $p_{c\ell}$). The efficiencies of the two tagging criteria on b jets ($\varepsilon_b^{Tag}$, $\varepsilon_b^{p_{Trel}}$) and on non-b jets ($\varepsilon_{c\ell}^{Tag}$, $\varepsilon_{c\ell}^{p_{Trel}}$) are also unknown, for a total of eight unknown quantities. Thus, a system of eight equations can be
Figure 13: Same as Fig. 12 using the IP distributions. The muon-jet $p_T$ is between 160 and 320 GeV.

written that relate the measurable quantities to the unknowns:

$$
\begin{align*}
    n & = n_b + n_{\ell f} \\
    p & = p_b + p_{\ell f} \\
    n^{\text{tag}} & = \varepsilon_b n_b + \varepsilon_{\ell f} n_{\ell f} \\
    p^{\text{tag}} & = \beta^{\ell g} \varepsilon_b p_b + \alpha^{\text{tag}} \varepsilon_{\ell f} p_{\ell f} \\
    n^{P_{\text{Trel}}} & = \varepsilon_{P_{\text{Trel}}} n_b + \varepsilon_{\ell f} n_{\ell f} \\
    p^{P_{\text{Trel}}} & = \beta^{P_{\text{Trel}}} \varepsilon_{P_{\text{Trel}}} p_b + \alpha^{P_{\text{Trel}}} \varepsilon_{\ell f} p_{\ell f} \\
    n^{\text{tag},P_{\text{Trel}}} & = \beta^n \varepsilon_b \varepsilon_{P_{\text{Trel}}} n_b + \alpha^n \varepsilon_{\ell f} \varepsilon_{P_{\text{Trel}}} n_{\ell f} \\
    p^{\text{tag},P_{\text{Trel}}} & = \beta^p \varepsilon_b \varepsilon_{P_{\text{Trel}}} p_b + \alpha^p \varepsilon_{\ell f} \varepsilon_{P_{\text{Trel}}} p_{\ell f} .
\end{align*}
$$

(3)

The method assumes that the efficiencies for a combination of tagging criteria are factorisable. Thus eight correlation factors are introduced to solve the system of equations: $\alpha^{\text{tag}}, \beta^{\text{tag}}, \alpha^{P_{\text{Trel}}}, \beta^{P_{\text{Trel}}}, \alpha^n, \beta^n, \alpha^p$, and $\beta^p$. These factors are obtained from the Monte Carlo simulation as a function of the muon-jet $p_T$ and $|\eta|$. The factors $\alpha$ and $\beta$ are determined for light-flavour and b jets, respectively. The superscripts $n$ and $p$ refer to the correlation between the two tagging efficiencies ($t^{\text{tag}}, p^{P_{\text{Trel}}}$) in the $n$ and $p$ samples, while the superscripts $t^{\text{tag}}$ and $p^{P_{\text{Trel}}}$ indicate the efficiency ratio between the $p$ and $n$ sample for the lifetime and $P_{\text{Trel}}$ criteria.

The simulation predicts that the correlation coefficients typically range between 0.95 and 1.05 for those associated with the b-tagging efficiencies, and between 0.7 and 1.2 for those associated...
4.3 Efficiency measurement using a reference lifetime tagger

While muon $p_{T\text{rel}}$ provides less distinction between jet flavours at large jet $p_T$, lifetime-based taggers retain their sensitivity to distinguish different jet flavours. In particular, the discriminant for the jet probability tagger has different distributions for different jet flavours at any value of jet $p_T$. The JP tagger also has the virtue of being calibrated directly in the data, where tracks with negative impact parameter can be used to compute the probability for those tracks to come from the primary vertex. The same calibration can be performed separately in simulated samples. The JP tagger can then serve as a reference tagger for estimating the fraction of b jets in a data sample, and also for estimating the fraction of b jets in a subsample that has been selected by an independent tagging algorithm. This allows a measurement of the efficiency of the independent algorithm. This lifetime tagger (“LT”) method can be performed in both inclusive-jet and muon-jet samples, and the results can be compared to obtain an estimate of the systematic uncertainty.

The efficiency measurement is performed in inclusive jet events in which at least one jet must be above a given $p_T$ threshold, and separately in dijet events in which at least one jet is a muon jet. To increase the fraction of b jets in the inclusive sample, an additional jet tagged by the JPM algorithm is also required. The sample with muon jets is already sufficiently enriched in b jets by the muon requirement. The same set of samples can be established with simulated events, so that the tagging efficiency can be measured there and a scale factor computed.

Because a value of the JP discriminant can be defined for jets that have as few as one track with a positive impact parameter significance, the discriminant can be calculated for most b jets, regardless of their $p_T$. The fraction of b jets that have JP information, $C_b$, rises from about 0.91 at $p_T = 20$ GeV to more than 0.98 for $p_T > 50$ GeV.

Figure 14 shows the JP discriminant distributions in the muon-jet sample and the inclusive sample, both before and after tagging the jets with an independent tagger, in this case the CSVM tagger. Also shown is the result of a fit to the distributions using JP-discriminant templates derived from simulations of b, c and light-flavour jets. The normalization of the relative flavour fractions $f_b$, $f_c$ and $f_{\text{light}}$ is left free, with the constraint that $f_b + f_c + f_{\text{light}} = 1$. The b-tagging efficiency can then be computed as the ratio of the number of b jets derived from the fit after and before applying the independent tagger, after correcting for the fraction of jets that have JP information:

$$\varepsilon_{b}^{\text{tag}} = \frac{C_b \cdot f_{b}^{\text{tag}} \cdot N_{\text{tag}}^{\text{data}}}{f_{b}^{\text{before tag}} \cdot N_{\text{data}}^{\text{before tag}}}.$$  \hspace{1cm} (4)

Examples of the efficiencies measured for the JPL and CSVM taggers are shown in Fig. 15. In both cases the results on simulation are close to those obtained from data.

This technique cannot be used to measure the efficiency of the JP tagger itself, as the JP discriminant is used in the fit to determine the b-jet content of the sample. However, the CSV discriminant, which is mostly based on information from secondary vertices, can be used in its place to determine the flavour content. More than 90% of jets have CSV information, as is the case with the JP discriminant. But unlike the JP discriminant, the CSV discriminant cannot be calibrated solely with the data. To remedy this, the CSV discriminant is used to estimate the tagging efficiency of the TC algorithms. By comparing these results to those using the JP
Figure 14: Fit to the JP discriminant with (top) a muon jet and (bottom) an inclusive jet with 260 < \( p_T \) < 320 GeV, (left) before and (right) after b-tagging with CSVM. Overflows are displayed in the upper right bin.
Figure 15: Efficiencies measured for the JPL (top) and CSVM (bottom) taggers with the LT method in the muon-jet sample. Closed and open symbols correspond to data and simulation, respectively.
discriminant, the bias due to using the CSV discriminant is determined to be (0-2%, 4-6%, 6-9%) for the (loose, medium, tight) operating points. The efficiencies and scale factors for the JP tagger are corrected for these biases.

4.4 Systematic uncertainties on efficiency measurements

Several systematic uncertainties affect the measurement of the b-tagging efficiency. Some are common to all four methods (PtRel, IP3D, System8, LT), some are common to a subset of them, and some are unique to a particular method.

Common systematic uncertainties for all methods:

- **Pile-up**: The measured b-tagging efficiency depends on the number of pp events superimposed on the primary interaction of interest. This number of pile-up events depends on the intensity of the proton beams and on the time interval between proton bunches (50 ns). In order to match the data better, each Monte Carlo event is reweighted in order to agree with the expected pile-up distribution. The systematic uncertainty is computed by varying the average value of the pile-up in data by 10% and calculating the difference in the values of $S_{fb}$ after reweighting the simulation with the modified distribution.

- **Gluon splitting**: Studies of angular correlation between b-jets at LHC [24] indicate that QCD events may have a larger fraction of gluon splitting into b\bar{b} pair than what is generated in the simulation. A study was carried out with the Monte Carlo sample where the number of events with gluon splitting was artificially changed by 50%. Results obtained with this modified gluon-splitting MC sample are then compared to those with the original sample. The observed deviation is quoted as a systematic uncertainty.

- **Muon $p_\mu^T$**: The central value of the b-tagging efficiency is extracted from data with muon $p_\mu^T > 5$ GeV. The choice of the selection cut affects the shape of the template distributions used in fits, and also the number of events used to measure the tagging efficiencies. The $p_\mu^T$ threshold is varied up to 9 GeV to test the sensitivity to this choice.

Common to the PtRel, IP3D and System8 methods:

- **Away jet tagger**: The dependency of the calculated b-tagging efficiency on the away jet tagger is studied by comparing the results obtained from tagging the away jets with different variants of the TC algorithm used for the central measurement (TCHEL, TCHEM, TCHPM). As the measured $S_{fb}$ tends to increase when the away tag is tighter, the maximum deviation from the default away jet tagger is taken as a systematic uncertainty.

Uncertainties unique to the PtRel method:

- **Shape of the light flavour $p_{Trel}$ spectrum**: The shape of the $p_{Trel}$ spectrum for light jets has been obtained by a data-driven technique, which minimizes the bias due to a mismodeling of the muon kinematics in the simulation. However, since the $p_{Trel}$ distribution in data is fit with a sum of templates for b jets and for c+udsg jets, uncertainties on the ratio between light and charmed jets in the simulation must be considered. To do so, the predicted ratio is varied by ±20%, and the fit is repeated, taking the variation in the results as a systematic uncertainty.

Uncertainties unique to the System8 method:
4.5 Efficiency measurement results

- **$p_{T_{\text{rel}}}$ selection cut**: One of the System8 criteria is a selection on the muon $p_{T_{\text{rel}}}>0.8$ GeV. In order to test the sensitivity to the b purity in the muon-jet sample and the relative charm/light fraction in the non-b background, this selection was changed from 0.5 to 1.2 GeV in the data. The correlation factors were recomputed accordingly in the simulation and the System8 method was applied again to the data in order to compute a b-tagging efficiency. The largest deviation observed from the central value is quoted as a systematic uncertainty.

- **Monte Carlo closure test**: The b-tagging efficiency can be directly calculated from the simulated QCD muon-enriched sample, as the flavour of the jets at generator level is known. In this case, the efficiency can be measured by taking the number of identified true b jets over all true b jets. The resulting value is denoted as the Monte Carlo truth b-tagging efficiency. The System8 method is also applied to this Monte Carlo sample. The resulting b-tagging efficiencies are in good agreement with the Monte Carlo truth, giving a negligible systematic uncertainty. (This systematic uncertainty does not appear for the other methods as they rely on template fits, making such a test trivial.)

Uncertainties unique to the LT method:

- **Fraction of b jets with JP information**: The fraction of inclusive jets with JP information is well described by the simulation. As explained above, the number of b jets before tagging is measured by a fit to the JP distribution and corrected by the fraction $C_b$ of b jets with JP information. A systematic uncertainty of half the residual correction, $(1-C_b)/(2C_b)$, is estimated from the simulation as a function of the b-jet $p_T$. A corresponding factor with a similar uncertainty is needed for measuring the efficiency of the JP and JBP taggers with the CSV discriminator spectrum.

- **Difference between muon jets and inclusive jets**: The difference between the measured $S_{F_b}$ in muon jets and in inclusive b jets is taken as a systematic uncertainty. This is the largest contribution to the systematic uncertainty on $S_{F_b}$ with the LT method. Due to the large statistical error on $S_{F_b}$ for inclusive jets with $p_T<80$ GeV, the same systematic uncertainty is used for $p_T<80$ GeV and within 80-210 GeV. If the difference on $S_{F_b}$ between muon jets and inclusive jets is smaller than the statistical error on $S_{F_b}$ for inclusive jets, this statistical error is used for the systematic uncertainty estimate.

- **Bias for the JP and JBP taggers**: The uncertainty on the measurement of the bias for using the CSV discriminant to measure the efficiency of the JP and JBP taggers as estimated for the TC taggers is propagated into the uncertainty on the scale factors for these taggers.

As an illustration, the breakdown of the systematic uncertainties on the data/MC scale factors of different tagging criteria is detailed in Tables 1-4 for the PtRel and System8 methods at low jet $p_T$ ($80 < p_T < 120$ GeV) and for the IP3D and LT methods at higher jet $p_T$ ($160 < p_T < 320$ GeV). In these momentum ranges the average uncertainty is about 3% for the PtRel method, 6-10% for the System8, 3-4%, for the IP3D method, and 2-7% for the LT method.

4.5 Efficiency measurement results

The methods described in the previous subsections cover a large range of jet $p_T$. Two of them, the PtRel and the System8 methods, provide precise measurements for the lower part of the spectrum while the other two, the IP3D and the LT methods, have been designed for high jet $p_T$. The measured data/MC scale factors are given in the first columns of Tab. 5 for low jet $p_T$ and
Table 1: Relative systematic uncertainties on $S_{F_b}$ for the PtRel method in the muon-jet $p_T$ range 80-120 GeV.

<table>
<thead>
<tr>
<th>b tagger</th>
<th>pile-up</th>
<th>$g \rightarrow b\bar{b}$</th>
<th>$p_T^\mu$</th>
<th>away jet</th>
<th>light $p_{Trel}$</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>1.6%</td>
<td>1.6%</td>
<td>1.1%</td>
<td>1.9%</td>
<td>0.4%</td>
<td>3.2%</td>
</tr>
<tr>
<td>JBPB</td>
<td>0.9%</td>
<td>0.5%</td>
<td>1.8%</td>
<td>1.2%</td>
<td>0.5%</td>
<td>2.5%</td>
</tr>
<tr>
<td>TCHEM</td>
<td>1.5%</td>
<td>0.5%</td>
<td>1.4%</td>
<td>1.6%</td>
<td>0.4%</td>
<td>2.7%</td>
</tr>
<tr>
<td>TCHPM</td>
<td>1.1%</td>
<td>0.1%</td>
<td>2.3%</td>
<td>1.1%</td>
<td>0.6%</td>
<td>2.8%</td>
</tr>
<tr>
<td>SSVHEM</td>
<td>0.7%</td>
<td>1.3%</td>
<td>2.0%</td>
<td>0.4%</td>
<td>0.4%</td>
<td>2.5%</td>
</tr>
<tr>
<td>CSVM</td>
<td>1.4%</td>
<td>1.3%</td>
<td>1.3%</td>
<td>1.2%</td>
<td>0.5%</td>
<td>2.6%</td>
</tr>
</tbody>
</table>

Table 2: Relative systematic uncertainties on $S_{F_b}$ for the System8 method in the muon-jet $p_T$ range 80-120 GeV.

<table>
<thead>
<tr>
<th>b tagger</th>
<th>pile-up</th>
<th>$g \rightarrow b\bar{b}$</th>
<th>$p_T^\mu$</th>
<th>away jet</th>
<th>$p_{Trel}$</th>
<th>MC closure</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>1.4%</td>
<td>0.6%</td>
<td>4.2%</td>
<td>3.9%</td>
<td>1.6%</td>
<td>0.1%</td>
<td>6.1%</td>
</tr>
<tr>
<td>JBPB</td>
<td>1.5%</td>
<td>1.9%</td>
<td>6.5%</td>
<td>1.5%</td>
<td>4.0%</td>
<td>0.0%</td>
<td>8.2%</td>
</tr>
<tr>
<td>TCHEM</td>
<td>1.3%</td>
<td>1.3%</td>
<td>6.6%</td>
<td>2.1%</td>
<td>2.4%</td>
<td>0.0%</td>
<td>7.5%</td>
</tr>
<tr>
<td>TCHPM</td>
<td>1.3%</td>
<td>2.7%</td>
<td>8.2%</td>
<td>1.9%</td>
<td>4.0%</td>
<td>0.1%</td>
<td>9.7%</td>
</tr>
<tr>
<td>SSVHEM</td>
<td>1.3%</td>
<td>0.1%</td>
<td>3.7%</td>
<td>2.8%</td>
<td>3.0%</td>
<td>0.0%</td>
<td>5.6%</td>
</tr>
<tr>
<td>CSVM</td>
<td>1.5%</td>
<td>0.4%</td>
<td>4.3%</td>
<td>1.3%</td>
<td>4.5%</td>
<td>0.1%</td>
<td>6.5%</td>
</tr>
</tbody>
</table>

Table 3: Relative systematic uncertainties on $S_{F_b}$ for the IP3D method in the muon-jet $p_T$ range 160-320 GeV.

<table>
<thead>
<tr>
<th>b tagger</th>
<th>pile-up</th>
<th>$g \rightarrow b\bar{b}$</th>
<th>$p_T^\mu$</th>
<th>away jet</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>0.1%</td>
<td>0.7%</td>
<td>0.1%</td>
<td>3.2%</td>
<td>3.2%</td>
</tr>
<tr>
<td>JBPB</td>
<td>0.2%</td>
<td>1.7%</td>
<td>0.4%</td>
<td>2.8%</td>
<td>3.3%</td>
</tr>
<tr>
<td>TCHEM</td>
<td>0.1%</td>
<td>1.2%</td>
<td>0.4%</td>
<td>2.4%</td>
<td>2.7%</td>
</tr>
<tr>
<td>TCHPM</td>
<td>0.2%</td>
<td>2.3%</td>
<td>0.7%</td>
<td>2.2%</td>
<td>3.3%</td>
</tr>
<tr>
<td>SSVHEM</td>
<td>0.6%</td>
<td>2.2%</td>
<td>0.3%</td>
<td>2.9%</td>
<td>3.6%</td>
</tr>
<tr>
<td>CSVM</td>
<td>0.6%</td>
<td>2.3%</td>
<td>0.1%</td>
<td>3.2%</td>
<td>4.0%</td>
</tr>
</tbody>
</table>

of Tab. 6 for high jet $p_T$. In these ranges the methods generally give compatible results within the quoted uncertainties. While some of the methods measure the efficiencies and scale factors only for muon jets, and not inclusive $b$ jets, simulation studies have shown that the difference in tagging efficiencies between the two are only a few percent. We assume that these small differences have no significant effect on the scale factors, which are relative data-to-simulation measurements. This assumption is supported by measurements in both samples using the JP method which show differences smaller than 5%.

The individual results have been combined to provide an optimal measurement of the data-to-simulation scale factors for $30 < p_T < 670$ GeV. For each jet $p_T$ range the most precise results have been used: the PtRel and System8 methods for $p_T < 120$, the IP3D method for $p_T > 120$ GeV and the LT method for the full momentum range.

The combination is based on a weighted mean of the scale factors in each bin [25]. The methods share an important part of their data sample (QCD di- and multijet events with at least one
4.5 Efficiency measurement results

Table 4: Relative systematic uncertainty on $S_{Fb}$ with the LT method in the muon-jet $p_T$ range 160-320 GeV.

<table>
<thead>
<tr>
<th>b tagger</th>
<th>pile-up</th>
<th>$g \rightarrow b\bar{b}$</th>
<th>$p_T^{\mu}$</th>
<th>$C_b$</th>
<th>inc. jets</th>
<th>bias</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>0.1%</td>
<td>0.8%</td>
<td>0.5%</td>
<td>0.1%</td>
<td>4.4%</td>
<td>4.0%</td>
<td>6.0%</td>
</tr>
<tr>
<td>J BPM</td>
<td>0.1%</td>
<td>0.4%</td>
<td>0.8%</td>
<td>0.1%</td>
<td>4.3%</td>
<td>4.0%</td>
<td>5.9%</td>
</tr>
<tr>
<td>TCHEM</td>
<td>0.1%</td>
<td>1.6%</td>
<td>0.3%</td>
<td>0.2%</td>
<td>2.8%</td>
<td>—</td>
<td>3.2%</td>
</tr>
<tr>
<td>TCHPM</td>
<td>0.2%</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.2%</td>
<td>1.7%</td>
<td>—</td>
<td>1.9%</td>
</tr>
<tr>
<td>SSVHEM</td>
<td>0.1%</td>
<td>2.3%</td>
<td>0.8%</td>
<td>0.2%</td>
<td>6.6%</td>
<td>—</td>
<td>7.0%</td>
</tr>
<tr>
<td>CSVM</td>
<td>0.2%</td>
<td>2.3%</td>
<td>0.7%</td>
<td>0.2%</td>
<td>5.2%</td>
<td>—</td>
<td>5.7%</td>
</tr>
</tbody>
</table>

Table 5: Data/MC scale factors $S_{Fb}$ for the medium operating points as measured with the PtRel, System8 and JP methods and their combination. Results are given for jet $p_T$ between 80 and 120 GeV. The first uncertainty on $S_{Fb}$ is statistical and the second is systematic except for the combination, where the total uncertainty is quoted.

<table>
<thead>
<tr>
<th>b tagger</th>
<th>$S_{Fb}$ (PtRel)</th>
<th>$S_{Fb}$ (System8)</th>
<th>$S_{Fb}$ (JP)</th>
<th>$S_{Fb}$ (comb.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>0.90 ± 0.01 ± 0.03</td>
<td>0.91 ± 0.03 ± 0.06</td>
<td>0.99 ± 0.02 ± 0.05</td>
<td>0.92 ± 0.03</td>
</tr>
<tr>
<td>JBPM</td>
<td>0.92 ± 0.01 ± 0.02</td>
<td>0.94 ± 0.03 ± 0.08</td>
<td>0.99 ± 0.02 ± 0.05</td>
<td>0.92 ± 0.03</td>
</tr>
<tr>
<td>TCHEM</td>
<td>0.94 ± 0.01 ± 0.03</td>
<td>0.97 ± 0.03 ± 0.07</td>
<td>0.98 ± 0.01 ± 0.03</td>
<td>0.95 ± 0.02</td>
</tr>
<tr>
<td>TCHPM</td>
<td>0.95 ± 0.01 ± 0.03</td>
<td>0.93 ± 0.02 ± 0.09</td>
<td>0.97 ± 0.01 ± 0.02</td>
<td>0.96 ± 0.02</td>
</tr>
<tr>
<td>SSVHEM</td>
<td>0.93 ± 0.01 ± 0.02</td>
<td>0.91 ± 0.03 ± 0.05</td>
<td>0.97 ± 0.01 ± 0.02</td>
<td>0.95 ± 0.02</td>
</tr>
<tr>
<td>CSVM</td>
<td>0.93 ± 0.01 ± 0.02</td>
<td>0.95 ± 0.03 ± 0.06</td>
<td>0.97 ± 0.01 ± 0.03</td>
<td>0.95 ± 0.02</td>
</tr>
</tbody>
</table>

Several sources of systematic uncertainties are also in common for all methods: the effects due to pile-up, gluon splitting and the selection criteria for muons. The muon $p_{Trel}$ and IP methods share the sensitivity to the choice of the away-jet tagger. The corresponding uncertainties were assumed to be fully correlated or anti-correlated according to the sign of the variations observed for the different methods. All other systematic effects are specific to individual methods and have been treated as uncorrelated. If the value of the $\chi^2$ obtained during the combination exceeded the number of degrees of freedom ($ndf$) the uncertainty on the combined value was

Table 6: Data/MC scale factors $S_{Fb}$ for medium operating points as measured with the IP3D and JP methods and their combination. Results are given for jet $p_T$ between 160 and 320 GeV. The first uncertainty on $S_{Fb}$ is statistical and the second is systematic except for the combination, where the total uncertainty is quoted.

<table>
<thead>
<tr>
<th>b tagger</th>
<th>$S_{Fb}$ (IP3D)</th>
<th>$S_{Fb}$ (JP)</th>
<th>$S_{Fb}$ (comb.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>0.93 ± 0.02 ± 0.03</td>
<td>0.99 ± 0.01 ± 0.06</td>
<td>0.95 ± 0.04</td>
</tr>
<tr>
<td>JBPM</td>
<td>0.97 ± 0.02 ± 0.03</td>
<td>0.99 ± 0.01 ± 0.06</td>
<td>0.97 ± 0.04</td>
</tr>
<tr>
<td>TCHEM</td>
<td>0.96 ± 0.02 ± 0.03</td>
<td>0.97 ± 0.01 ± 0.03</td>
<td>0.96 ± 0.03</td>
</tr>
<tr>
<td>TCHPM</td>
<td>0.97 ± 0.02 ± 0.03</td>
<td>0.97 ± 0.01 ± 0.02</td>
<td>0.97 ± 0.02</td>
</tr>
<tr>
<td>SSVHEM</td>
<td>0.99 ± 0.02 ± 0.04</td>
<td>0.98 ± 0.01 ± 0.07</td>
<td>0.98 ± 0.04</td>
</tr>
<tr>
<td>CSVM</td>
<td>0.95 ± 0.02 ± 0.04</td>
<td>0.97 ± 0.01 ± 0.06</td>
<td>0.96 ± 0.04</td>
</tr>
</tbody>
</table>
Figure 16: Individual and combined measurements of the ratio of efficiencies in data and simulation for the JPL (a) and CSVM (b) taggers. The upper panels show the individual measurements from the muon $p_T$ (“PtRel”), System8 (“System8”), muon IP (“IP3D”) and lifetime (“LT”) tagger methods. The thick lines indicate the statistical errors while the thin ones correspond to the combined statistical and systematic uncertainties. The hatched areas represent the combined measurements. In the lower panels, the thick lines indicate the statistical error while the thin lines correspond to the combined systematic and systematic uncertainties. The hatched area represents the combined measurements. The error bars attached to the function have the same size as the uncertainties from the combined measurements in each bin.

The error bars attached to the function have the same size as the uncertainties from the combined measurements in each bin. The upper panels show the individual measurements from the muon $p_T$ (“PtRel”), System8 (“System8”), muon IP (“IP3D”) and lifetime (“LT”) tagger methods. The thick lines indicate the statistical errors while the thin ones correspond to the combined statistical and systematic uncertainties. The hatched areas represent the combined measurements. In the lower panels, the thick lines indicate the statistical error while the thin lines correspond to the combined systematic and systematic uncertainties. The error bars attached to the function have the same size as the uncertainties from the combined measurements in each bin.
conservatively rescaled by \( \sqrt{\chi^2/ndf} \). Summaries for the individual and combined scale factor measurements for the JPL and the CSVM taggers are shown in Fig. 16 together with parameterizations of the combined SF of the form \( SF_b(p_T) = a(1 + \beta p_T)/(1 + \gamma p_T) \). Combined values for a low and a high jet \( p_T \) range are shown in the last columns of Tab. 5 and 6, respectively.

The combined values have been compared to the results obtained in studies of top quark pair production [5]. In order to allow for a comparison the \( p_T \) dependent scale factors measured in di- and multijet events have been reweighted to match the jet-\( p_T \) spectrum observed in \( t\bar{t} \) events. The results from both types of events are shown in Tab. 7 for the medium operating points. Excellent agreement is observed.

Table 7: Values and total uncertainties for the efficiency scale factors \( SF_b \) obtained in multijet and \( t\bar{t} \) events for \( b \) jets in the expected \( p_T \) range of \( t\bar{t} \) events. For the \( t\bar{t} \) results with the JP and JBP algorithms the profile likelihood ratio values [5] are quoted as they correspond to the same calibration as for the multijet results.

<table>
<thead>
<tr>
<th>b tagger</th>
<th>( SF_b ) in multijet events</th>
<th>( SF_b ) in ( t\bar{t} ) events</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>0.92 ± 0.03</td>
<td>0.95 ± 0.03</td>
</tr>
<tr>
<td>JBPM</td>
<td>0.92 ± 0.03</td>
<td>0.93 ± 0.04</td>
</tr>
<tr>
<td>TCHEM</td>
<td>0.95 ± 0.03</td>
<td>0.96 ± 0.04</td>
</tr>
<tr>
<td>TCHPM</td>
<td>0.94 ± 0.03</td>
<td>0.93 ± 0.04</td>
</tr>
<tr>
<td>SSVHEM</td>
<td>0.95 ± 0.03</td>
<td>0.96 ± 0.04</td>
</tr>
<tr>
<td>CSVM</td>
<td>0.95 ± 0.03</td>
<td>0.97 ± 0.04</td>
</tr>
</tbody>
</table>

5 Measurement of the misidentification rate

The measurement of the mistagging rate from light-flavour jets relies on the definition of negative discriminator values for each \( b \)-tagging algorithm. These “negative taggers” can then be used in the same way as the regular \( b \)-tagging algorithms both in data and in the simulation. As the negative-tagged jets are enriched in light flavours, the mistagging rate can be measured from the data, with the simulation used to extract a correction factor.

The mistagging rate is evaluated from tracks with a negative impact parameter or from secondary vertices with a negative decay length (see Sect. 3). When a negative tagger is applied to jets of any flavour, the corresponding tagging efficiency is denoted “negative tag rate”. The negative and positive \( b \)-tagging discriminators distributions in data are compared with the simulation in Fig. 17. The events are selected by requiring jet triggers with a \( p_T \) threshold of 30 GeV, corresponding to an average \( p_T \) of 44 GeV. For all \( b \)-tagging algorithms, the data and simulation are found to be in agreement within about ±20%. Similar results are found for a sample of events selected by requiring jet triggers with a \( p_T \) threshold of 300 GeV, in which the average \( p_T \) is 213 GeV.

The mistagging rate is evaluated as:

\[
\varepsilon_{\text{data}}^{\text{mistag}} = \varepsilon_{\text{data}}^{-} \cdot R_{\text{light}}.
\]

\( \varepsilon_{\text{data}}^{-} \) is the negative tag rate as measured in jet data, defined as the fraction of jets that are negatively tagged. \( R_{\text{light}} = \varepsilon_{\text{MC}}^{\text{mistag}} / \varepsilon_{\text{MC}}^{-} \), a correction factor taken from simulation, is the ratio of the mistagging rate for light-flavour jets to the negative tag rate for jets of all flavours in the simulation.
Figure 17: Signed b-tag discriminators in data (dots) and simulation for light-flavour jets (blue area, with a lighter colour for the negative discriminators), c jets (green area) and b jets (red area) for the TCHE, JP, SSVHE and CSV algorithms (from top left to bottom right). A jet-trigger $p_T$ threshold of 30 GeV is required for both data and simulation. The simulation is normalised to the number of entries in the data. Underflow and overflow entries are displayed in the lower and upper bins, respectively.
\( \varepsilon_{\text{data}} \) depends on the fractions of c and b quarks in the negatively-tagged jets (which tend to decrease \( R_{\text{light}} \)), on residual differences between light quark flavour and gluon jets, and on the fractions of tracks from other displaced processes such as \( K_S^0 \) and \( \Lambda \) decays, interactions in the detector material and mismeasured tracks (which tend to increase \( R_{\text{light}} \)). Due to these various contributions, the Monte Carlo simulation predicts ranges of \( R_{\text{light}} \) for the different algorithms and jet \( p_T \) values of about 1.1 to 1.4 (1 to 2, 1 to 4) for the loose (medium, tight) operating points, respectively.

To connect the measured mistagging rate to that predicted by the simulation, a scale factor \( S_{\text{light}} \) is defined:

\[
S_{\text{light}} = \frac{\varepsilon_{\text{data}}}{\varepsilon_{\text{MC}}}.
\]

(6)

The following systematic effects on the mistagging rate based on negative tags are considered:

- **b and c fractions**: the fraction of b-flavour jets has been measured in CMS to agree with the simulation within a \( \pm 20\% \) uncertainty [26]. A \( \pm 20\% \) uncertainty is conservatively estimated for the overall fraction of b and c jets. The b+c flavour fraction is varied in the QCD Monte Carlo, from which a systematic uncertainty on \( R_{\text{light}} \) is inferred.

- **Gluon fraction**: this affects both the mistagging rate in simulation and the overall negative tag rates. The average fraction of gluon jets depends on the details of the parton density and hadronisation functions used in the simulation. An uncertainty of \( \pm 20\% \) is extracted from the comparison of simulation with data [27].

- **Long lived \( K_S^0 \) and \( \Lambda \) decays**: the amount of reconstructed \( K_S^0 \) and \( \Lambda \) are found to be a factor 1.40 \( \pm 0.15 \) and 1.50 \( \pm 0.50 \) larger, respectively, in the data than in the simulation, with the quoted uncertainty accounting for the \( p_T \) dependence [28]. To estimate the uncertainty on \( R_{\text{light}} \) due to the \( K_S^0 \) and \( \Lambda \) contribution, the simulated QCD events are reweighted in order to match the observed yield of \( K_S^0 \) and \( \Lambda \) in the data. Then this yield is varied within the quoted uncertainty and the inferred variation on \( R_{\text{light}} \) taken as a systematic uncertainty.

- **Photon conversion and nuclear interactions**: the rate of secondary interactions in the pixel detector layers has been measured with \( \pm 5\% \) precision [29, 30]. The corresponding variation implies a systematic uncertainty on \( R_{\text{light}} \).

- **Mismeasured tracks**: according to the simulation, jets with a reconstructed track not associated with a genuine charged particle also present an excess of positive over negative tags. To correct for residual mismeasurement effects, a \( \pm 50\% \) variation on this contribution is taken into account in the systematic uncertainty on \( R_{\text{light}} \).

- **Sign flip**: The ratio of the number of negative over positive tagged jets is computed in a muon-jet sample similar to the one described in Sect. 4, with a larger than 80\% b purity. Data and simulation are found to be in good agreement. From the statistical uncertainty on the comparison, the absolute uncertainty on this ratio is estimated as 2\% (1\%, 0.5\%) for loose (medium, tight) operating points, respectively. This sign flip uncertainty can be translated into a systematic uncertainty on \( R_{\text{light}} \).

- **Pile-up**: the mistagging rate depends on the pile-up model used in the simulation. The simulated events are reweighted in order to match the pile-up rate in the data. Differences between \( R_{\text{light}} \) values obtained for different running periods are used to estimate the systematic uncertainty, which is about \( \pm 1\% \) for all taggers.

- **Event sample**: Physics analyses use jets from different event topologies. For a given
jet $p_T$, the mistagging rate is different if the jet is the leading one or if there are other jets with higher $p_T$ values in the same event. Measured mistagging scale factors for leading and sub-leading jets have a dispersion of about 7%. In addition, mistagging scale factors vary by 2-7%, depending on the tagger, for different running periods. These two uncertainties are added in quadrature to account for an uncertainty due to sample dependence. This is the dominant contribution to the overall systematic uncertainty on the mistagging rate.

The systematic uncertainties are detailed in Table 8 for the various algorithms and operating points in the jet $p_T$ range between 80 and 120 GeV.

### Table 8: Relative systematic uncertainties on $SF_{light}$ for jet $p_T$ in the range 80 - 120 GeV.

<table>
<thead>
<tr>
<th>b tagger</th>
<th>c+jets</th>
<th>gluon</th>
<th>$V^0$ + 2$^\text{nd}$ int.</th>
<th>mis-meas.</th>
<th>sign flip</th>
<th>MC stat</th>
<th>sample + pile-up</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>6.2%</td>
<td>1.2%</td>
<td>3.9%</td>
<td>0.5%</td>
<td>1.6%</td>
<td>0.9%</td>
<td>9.0%</td>
<td>11.9%</td>
</tr>
<tr>
<td>JBPM</td>
<td>4.5%</td>
<td>0.8%</td>
<td>3.6%</td>
<td>1.2%</td>
<td>5.1%</td>
<td>0.7%</td>
<td>8.0%</td>
<td>11.3%</td>
</tr>
<tr>
<td>TCHEM</td>
<td>1.6%</td>
<td>1.0%</td>
<td>1.8%</td>
<td>0.6%</td>
<td>2.5%</td>
<td>0.6%</td>
<td>9.2%</td>
<td>10.0%</td>
</tr>
<tr>
<td>TCHPM</td>
<td>1.0%</td>
<td>0.9%</td>
<td>2.4%</td>
<td>1.9%</td>
<td>2.9%</td>
<td>0.7%</td>
<td>7.3%</td>
<td>8.7%</td>
</tr>
<tr>
<td>SSVHEM</td>
<td>3.2%</td>
<td>1.8%</td>
<td>3.0%</td>
<td>0.7%</td>
<td>4.6%</td>
<td>0.7%</td>
<td>7.4%</td>
<td>10.1%</td>
</tr>
<tr>
<td>CSVM</td>
<td>3.2%</td>
<td>1.8%</td>
<td>3.0%</td>
<td>0.7%</td>
<td>4.6%</td>
<td>0.7%</td>
<td>7.4%</td>
<td>10.1%</td>
</tr>
</tbody>
</table>

Table 9: Mistagging rate and data-to-simulation scale factor for the medium operating points for jet $p_T$ in the range 80 - 120 GeV. The statistical errors are quoted for the mistagging rates and the statistical+systematic uncertainties for the scale factors.

<table>
<thead>
<tr>
<th>b tagger</th>
<th>mistag rate ($\pm$stat)</th>
<th>scale factor ($\pm$stat $\pm$ syst)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPM</td>
<td>0.0109 $\pm$ 0.0002</td>
<td>1.02 $\pm$ 0.02 $\pm$ 0.16</td>
</tr>
<tr>
<td>JBPM</td>
<td>0.0112 $\pm$ 0.0001</td>
<td>0.94 $\pm$ 0.01 $\pm$ 0.11</td>
</tr>
<tr>
<td>TCHEM</td>
<td>0.0286 $\pm$ 0.0003</td>
<td>1.20 $\pm$ 0.01 $\pm$ 0.14</td>
</tr>
<tr>
<td>TCHPM</td>
<td>0.0306 $\pm$ 0.0003</td>
<td>1.24 $\pm$ 0.01 $\pm$ 0.12</td>
</tr>
<tr>
<td>SSVHEM</td>
<td>0.0209 $\pm$ 0.0002</td>
<td>0.93 $\pm$ 0.01 $\pm$ 0.08</td>
</tr>
<tr>
<td>CSVM</td>
<td>0.0152 $\pm$ 0.0002</td>
<td>1.10 $\pm$ 0.01 $\pm$ 0.11</td>
</tr>
</tbody>
</table>

The measured mistagging rates and data/simulation scale factors are presented in Figs. 18-19 as a function of the jet $p_T$ and $|\eta|$ for the JPL and CSVM taggers. The observed scale factors are close to one over a broad range of $p_T$ and $|\eta|$. A jet-trigger $p_T$ threshold of 80 GeV is required for the jet $|\eta|$ distribution, corresponding to an average jet $p_T$ of 79 GeV in this sample.

As an illustration, the data mistagging rates and data/simulation scale factors are given in Tab. 9 for jets with $p_T$ between 80 and 120 GeV.

### 6 Conclusions

The CMS experiment has developed a variety of algorithms that can be used to identify jets that arise from the hadronisation and decay of bottom quarks. These algorithms exploit the kinematics of the hadron jets, and are robust against imperfections in the detector alignment, and against the presence of multiple p-p collisions per beam crossing.
Figure 18: For the JPL tagger: mistagging rate in data (red squares) and simulation (blue dots) (top); scale factor for the mistagging rate (bottom). The last $p_T$ bin in each plot includes all jets with $p_T > 520$ GeV. The solid curves are the results of polynomial fits to the data points. The dashed curves represent the overall statistical and systematic uncertainties on the measurements.

Figure 19: Same as Fig. 18 for the CSVM tagger.
The efficiency of the tagging algorithms has been measured through several different methods in samples of multijet events recorded by CMS. The results from all methods are consistent with each other over a wide range of jet transverse momenta, $30 < p_T < 670$ GeV. The measured efficiencies for b jets agree with those predicted by the CMS simulation within $\approx 10\%$. In the $p_T$ range $30 - 320$ GeV (320 - 670 GeV) and for a mistagging rate of about 1%, the uncertainties on the data-to-simulation scale factors are smaller than 5% (9%). The corresponding precision on the scale factors for the misidentification probabilities is about 9 - 16 % in the full $p_T$ range.

The performance of the b-tagging algorithms is a crucial element of measurements of the properties of known particles, and in the search for new particles, that decay to b quarks. The demonstrated quality of the algorithms will thus be a driver of the success of the CMS physics programme.

References


[17] ALEPH Collaboration, “A precise measurement of $\Gamma_{Z\rightarrow h\bar{h}} / \Gamma_{Z\rightarrow \text{hadrons}}$”, *Phys. Lett.* **B313** (1993) 535. doi:10.1016/0370-2693(93)90028-G.


