b-Jet Identification in the DØ Experiment

C.M. Anastasoaie [b,1], D. Bloch [a], D. Boline [e], B. Clément [a],
L. Feligioni [e,2], F. Filthaut [b], A. Garcia-Bellido [j], S. Grédèr [a],
H. Greenlee [d], A. Haas [h], A. Khanov [i], M. Narain [f],
I. Ripp-Baudot [a], F. Rizatdinova [i], T. Scanlon [c],
A. Schwartzman [g,3], V. Siccardi [a], G. Watts [j]

[a]*IPHC, IN2P3/CNRS, Strasbourg, France*

[b]*Radboud University Nijmegen and Nikhef, Nijmegen, The Netherlands*

[c]*Imperial College, London, United Kingdom*

[d]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*

[e]*Boston University, Boston, Massachusetts 02215, USA*

[f]*Brown University, Providence, Rhode Island 02912, USA*

[g]*Princeton University, Princeton, New Jersey 08544, USA*

[h]*Columbia University, New York, New York 10027, USA*

[i]*Oklahoma State University, Stillwater, Oklahoma 74078, USA*

[j]*University of Washington, Seattle, Washington 98195, USA*

Abstract

Key words: b-jet identification, b-tagging, DØ, Tevatron, Collider
PACS: 29.85.+c

1 Introduction

The bottom quark occupies a rather special place in our understanding of the fundamental fermions that are the constituents of matter: on the one hand, its mass (of the order of 5 GeV/c^2 [1]) is substantially larger than that of the (next heaviest) charm quark. On the other hand, it is sufficiently light to be produced copiously at high energy colliders.

In particular, unlike the top quark, the bottom quark is lighter than the W boson, preventing decays to on-shell W bosons. As a result, it lives long enough for hadronization to occur before its decay. The average lifetime of b-flavored hadrons (referred to as B hadrons in the following) has been measured to be about 1.5 ps [1]: this is sufficiently long for B hadrons, even for moderate momenta, to travel distances of the order of at least a mm. Combined with the relatively large mass of B hadrons, the use of precise tracking information therefore allows to detect the presence of B hadrons through their charged decay products. In addition, B hadron decays often lead to the production of high momentum leptons: especially in the QCD dominated final states at hadron colliders, the observation of such leptons provides easy access to samples with enhanced b-jet content. The identification of jets originating from the hadronization of bottom quarks (referred to as b-jet identification or b-tagging in the following) in the DØ experiment is the subject of this publication.

1.1 The upgraded DØ detector

The DØ experiment is one of the two experiments operating at the Tevatron $p\bar{p}$ Collider at Fermilab. After a successful Tevatron Run I, which led to the discovery of the top quark [2], the Tevatron was upgraded to provide both a higher center-of-mass energy (from 1.8 TeV to 1.96 TeV) and a significant increase in luminosity. Run II started in 2001, and the Tevatron has delivered 1.2 fb^{-1} of integrated luminosity to the experiments by March, 2006, at which time another luminosity upgrade was commissioned. This publication refers to the Run II data taken before March, 2006.

To cope with the increased luminosity and decreased bunch spacing (from 3.6 μs to 396 ns), also the DØ detector underwent a significant upgrade; the upgraded detector is described in detail elsewhere [3]. In particular, a 2T central solenoid was installed to provide an axial magnetic field used to measure the momentum of charged particles. Correspondingly, the existing tracking detectors were removed and replaced with two new detectors:

- the Central Fiber Tracker (CFT), consisting of about 77,000 axial and small-angle stereo scintillating fibers arranged in eight concentric layers;

- and the Silicon Microstrip Tracker (SMT), a detector featuring 912 silicon strip sensor modules arranged in six barrel and 16 disk structures. Of particular interest is the innermost layer of SMT barrel ladders: its proximity to the beam line (at a radius of 2.6 cm) results in a relatively small extrapolation to the beam line, and hence in good vertex reconstruction capabilities.

The ability to identify efficiently the b-quarks [4] in an event considerably broadens the range of physics issues that can be addressed by the DØ experiment in Run II. While the analysis leading to the observation of the top quark by DØ in Run I could employ only semimuonic decays $b \rightarrow \mu \bar{\nu}_\mu X$, the use of lifetime tagging allows for a more precise determination of the top quark properties (see *e.g.* [4]). Evidence for the production of top quarks through electroweak production processes has not yet been established, but upper limits constraining the cross sections for these processes approach their predictions made by the standard model (SM) of particle physics. The search for the SM Higgs boson also benefits from b-jet identification: a relatively light ($m_H \lesssim 135 \text{ GeV}/c^2$) Higgs boson will decay predominantly to $b\bar{b}$ quark pairs. Finally, many theoretical models invoking supersymmetry exhibit a strongly enhanced cross section for the production of final states containing b-quarks.

This article is subdivided as follows. Section 2 describes the objects that serve as input to the b-tagging algorithms. Section 3 introduces clean-up criteria applied before the tagging proper. Sections 4, 5, and 6 describe the basic ways in which lifetime-correlated variables are extracted. Section 7 combines these variables in an artificial neural network to obtain an optimal tagging performance. Finally, Sections 8 and 9 detail how collider data are used to calibrate the performance of the resulting tagging algorithm.

---

[4] In this article, charge conjugated states are implied as well.

2 Object reconstruction

The input for lifetime identification of b quark jets consists of three kinds of reconstructed objects.

- Charged particle tracks, which are reconstructed from hits in the CFT and SMT tracking detectors.
- Primary and secondary vertices, which are built from two or more charged particle tracks that originate from a common point in space.
- Hadron jets, which are reconstructed primarily by their energy deposition in the calorimeter.

2.1 Charged particle tracking

Tracks are reconstructed using the CFT and SMT detectors. A diagram of the tracking detectors is shown in Fig. 1
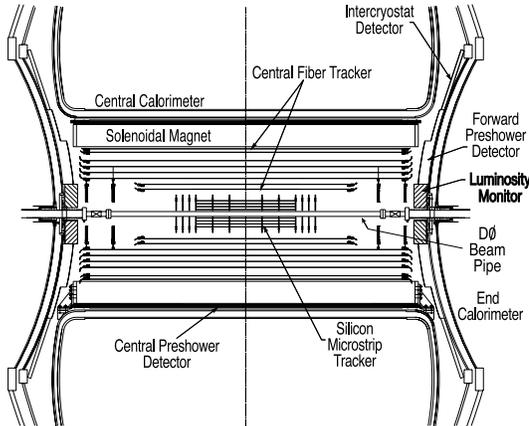
Fig. 1. Central Tracking Detectors.

In the DØ experiment, tracks are found by two independent algorithms, which are then combined, filtered, and refit. The two track-finding algorithms are called HTF (histogramming track finder) and AA (alternative algorithm). The HTF algorithm is based on a histogramming techique. The AA algorithm uses a road-following (Kalman Filter [5]) technique starting from seed tracks. The two algorithms are described in more detail below.

2.1.1 HTF Track Finding Algorithm

A general track is characterized by five track parameters, which in DØ are conventionally chosen to be $(d, z, \phi, \tan \lambda, \rho)$, where d is the beam line im-

pact parameter, z is the axial coordinate at the beam line, ϕ is the azimuthal direction, $\tan \lambda = dz/dr$ and ρ is the radius of curvature. Tracks that are produced by beam particle interactions invariably have small impact parameters ($d \approx 0$), so for the purpose of pattern recognition one can consider that tracks occupy a four dimensional parameter space. If we restrict attention to the axial view, the number of relevant track parameters reduces to two, namely, ρ and ϕ. Thus, axial tracks can be thought of as points in the (ρ, ϕ) plane.

Now consider axial hits, which are space points in the (x, y) plane. A single axial hit is consistent with a one-parameter family of axial tracks, which approximately makes a straight line in (ρ, ϕ) space. The mapping from axial space point (x, y) to a line in axial track parameter space (ρ, ϕ) is called a Hough Transform [6]. The pattern recognition problem then reduces to finding points in track parameter space (axial tracks) where lines corresponding to several hits intersect.

The pattern recognition problem is solved using a histogram to represent (ρ, ϕ) space. Axial hits are then added to the histogram in the roughly line of cells consistent with that hit. Tracks show up as spikes in the histogram. Cells containing a number of hits above a minimum threshold are candidate tracks.

Once axial track candiates have been found, stereo information can be added in the same way. Stereo tracks can be thought of as points in $(z, \tan \lambda)$ space. Stereo hits are points in (r, z) space that map via a Hough Transform to lines in $(z, \tan \lambda)$ space.

In practice, track candidates found by histogramming tend to have high combinatorial background. Therefore, following each histogramming pass, track candidates are filtered using a Kalman Filter track fitter.

The full HTF algorithm utilizes two independent passes to maximize efficiency. In the first pass, tracks are found in the SMT using the histogramming method described above, then extended into the CFT using a Kalman Filter. In the second pass, tracks are found in the CFT, then extended into the SMT.

2.1.2   AA Track Finding Algorithm

The AA algorithm differs from the HTF algorithm primarily in the way it generates seed tracks. Like HTF, AA uses two pattern recognition passes, SMT to CFT, and CFT to SMT.

In the first pass, seed tracks are constructed using all combinations of three 2D (axial + stereo) SMT hits that are consistent with a helical trajectory passing close to the beam line. Seed tracks are then extended to the remainder of the SMT and the CFT.

In the second pass, seed tracks are constructed using all combinations of three 2D (axial + stereo) CFT hits. For the second pass, seed tracks are required to pass close to a primary vertex found using tracks from pass 1 (not just close to the beam line). Seed tracks are then extended to the remainder of the CFT and the SMT.

2.1.3 Track Filtering and Fitting

All phases of track reconstruction make extensive use of the Kalman Filter [5] algorithm for reducing fake tracks and obtaining estimates of track parameters. A final Kalman Filter track fit is done after HTF and AA pattern recognition is finished. The final pass differs from earlier passes in that it uses a more sophisticated model of track propagation that is optimized for accuracy rather than speed.

All Kalman Filter type tracking algorithms share certain features. The Kalman Filter requires a partially reconstructed candidate track or seed track based on a subset of avilable tracking detectors to get started. The candidate track is propagated to the next detector surface, giving a predicted position and error on that surface. Hits on the detector surface near the predicted track can be added to the track to give updated estimates of track parameters, and an incremental chisquare. Hits that produce too large of an incremental chisquare are rejected. It can happen that more than one hit, or no hit, is consistent with the track prediction. Tracks are extended in this way until no more extension is possible. In addition to being rejected by an incremental chisquare cut, tracks can be rejected at any step, or after the final step, by global properties, but especially the total the total number and types of hits and misses. As a final step, tracks that share too many hits are filtered.

It would scarcely be possible to enumerate all cuts used in track reconstruction due to high complexity. However, here is a partial list of cuts that is used in DØtrack reconstruction.

List tracking cuts.

2.1.4 Tracking Acceptance and Performance

2.2 Primary vertex reconstruction

The reconstruction and identification of the primary vertex at DØ consists of the following steps: (i) track selection; (ii) vertex fitting by means of an adaptive algorithm; and (iii) primary vertex selection.

In the first stage, tracks are selected if they have $p_T > 0.5$ GeV and two or more hits in the silicon detector if the track is within the geometric acceptance of the silicon detector as measured in the (η, z) plane. The selected tracks are then clustered along the z-direction in 2 cm regions to separate groups of tracks coming from different interactions.

The second stage involves the fitting of tracks to a vertex on each of the z clusters obtained before. This is done in two passes. In the first pass, the location and width of the beam in the transverse plane of each event is determined by fitting the selected tracks in each cluster to a common vertex. A Kalman Filter vertex fitter is used for this step where tracks with the highest χ^2 contribution to the vertex are removed in turn, until the total vertex χ^2 per degree of freedom is smaller than 10. In the second pass, we refine the track selection in each z cluster by keeping only those tracks whose distance of closest approach (d) significance to the beam spot location is less than 5. To determine the d/σ_d value of each track, the beam spot location and its error as determined in the first pass is used. Thus the resolution of the beam spot determination in the first pass is included in the second pass. Once the outliers with respect to the beam position have been removed from the selected tracks, an adaptive vertex algorithm [7] is used to fit the selected tracks into a common vertex in each cluster.

The adaptive algorithm is specially suited to reduce the contribution of distant tracks to the vertex fit, thus obtaining a better separation between primary and secondary vertices. This algorithm is an improvement over the Kalman Filter vertex fitter in that all tracks are allowed to contribute to the final vertex fit instead of rejecting those tracks whose χ^2 contribution to the vertex fit is larger than some value. In this algorithm, each track is given a weight depending on its χ^2 contribution to the fitted vertex, and this is done iteratively.

The final and third step consists on selecting which of the fitted vertices in all z clusters is the result of the hard scatter interaction. The hard scatter vertex is distinguished from other soft-interaction vertices by the higher average p_T of its tracks. We form a probability function for each vertex based on the p_T of tracks originated in minimum bias vertices, and then select the vertex with the lowest probability.

The reconstruction and identification efficiency in data is between 97 and 100% for primary vertices reconstructed up to $z = 100$ cm. For multijet events, the position resolution of the primary vertex in the transverse plane is around 10 μm and in the longitudinal direction of around 25 μm.

We should also say something about the selection! And some plots would be welcome.

The vast majority of data analyses in DØ make use of so-called cone jets, which collect all calorimeter energy deposits within a fixed angular distance R. In particular, the cone jet reconstruction algorithm used within DØ is the *Run II cone jet* algorithm [8]. This algorithm is insensitive to the presence of soft or collinear radiation off partons, thus allowing for detailed comparisons of jet distributions in the DØ data with theoretical predictions. The cone radii used in analyses in DØ are $R = 0.5$ and $R = 0.7$; but for most high p_T physics only the $R = 0.5$ cone jets are used, and it is only these jets that are described in this article.

As the jets are reconstructed on the basis of calorimetric information, for such comparisons to be possible corrections are applied for various effects:

- energy deposits not from the hard interaction (either from the underlying event, the remnant of the original p$\bar{\text{p}}$ system, or from additional soft interactions);
- the (energy dependent) calorimeter response to incident high-energy particles;
- charged particles can be swept into or out of the cone on their path from the interaction point to the calorimeter.

This topic of the determination of the *jet energy scale* (JES) is described in a separate paper [**?**]. The resulting JES, by itself, does not yet account for missing neutrinos from decays of b- or c-flavoured hadrons. While such additional corrections may be important for physics analyses, b-tagging is only sensitive to it because the tagging performance obtained in data (Sect. 8) is parametrized in terms of jet E_{T} (and η) and applied to MC jets. For this purpose, corrections for undetected neutrinos (and for the energy not deposited in the calorimeter, in the case of muons associated with jets) are not applied.

3   Preliminaries

3.1   Signal Definition

As indicated in the preceding Section, the b-tagging algorithms used within the DØ experiment are jet based rather than event based. This choice makes sense especially for high-luminosity hadron colliders, where pile-up from previous interactions, as well as multiple interactions in the same bunch crossing, may lead to other tracks and jets in the event besides those of the "interesting" high p_T interaction.

However, this choice introduces an ambiguity for jets in simulated events. In order to estimate the performance of a b-tagging algorithm it is first necessary to specify precisely how a jet's flavour is determined. The following choice has been made:

- if at the particle level a B hadron is found within a $R = 0.5$ radius of the the jet direction, the jet is considered to be a b jet;
- if no B hadron but a C hadron is found, the jet is considered to be a c jet;
- if no C hadron is found either, the jet is considered to be a light flavour jet.

This choice is preferred over the association with a b or c quark, as in the latter case, parton showering may lead to a large distance between the original quark direction and that of the corresponding jet(s).

3.2   Taggability

The tagging algorithms described in the following Sections are based entirely on tracking and vertexing of charged particles. Therefore, a very basic requirement is that there should be charged particle tracks associated with the (calorimeter) jet. Rather than incorporating such basic requirements in the tagging algorithms themselves, they are implemented as a separate step. There are several advantages to a staged approach:

- as detailed in Section 8, the tagging algorithm's performance must be evaluated on real data. It is parametrized in terms of the jet kinematics (E_T and $|\eta|$). This parametrization presupposes that there are no further dependences. However, the interaction region at the DØ detector is quite long, $\sigma_z \approx 25$ cm, and the detector acceptance affects the track reconstruction efficiency dependence on η differently for different values of the interaction point's z coordinate;
- some of the reconstructed calorimeter jets result from electronics noise. Al-

though such fake jets are suppressed by calorimeter-based jet identification criteria, the association with tracks offers additional discrimination. The fraction of fake jets is generally small, **(quote a number here?)** but it depends on the final state under consideration. Decoupling this effect from the tagging algorithms proper allows to extract a tagging performance which can be assumed to be *universal*, *i.e.*, applicable to general final states.

The requirement for a jet to be *taggable*, *i.e.*, for it to be considered for further application of the tagging algorithms, is that it should be within a $R = 0.5$ distance from a so-called track jet. Track jets are reconstructed starting from tracks having at least one hit in the SMT, a distance to the selected primary vertex less than 2 mm in the transverse plane and less than 4 mm in the z direction, and $p_{\mathrm{T}} > 0.5$ GeV/c. Starting with "seed" tracks having $p_{\mathrm{T}} > 1$ GeV/c, the Snowmass Jet algorithm [9] is used to cluster the tracks within cones of radius $R = 0.5$. **Here I omit the z clustering. Given that the $z_{\mathbf{dca}}$ cut of 4 mm is used, that should be of no importance??**

Here should follow some representative plots on taggability.

3.3 V^0 Rejection

By construction, the lifetime tagging algorithms assume that any measurable lifetime is indicative of heavy flavour jets. However, also several strange hadrons (in particular, K_{S}, K_{L}, and Λ, commonly denoted as V^0s) decay weakly, with lifetimes of tens of ps (or even tens of ns in the case of the K_{L}). In additions, $\gamma \rightarrow \mathrm{e}^+\mathrm{e}^-$ conversions may occur in the detector material at large distances from the beam line.

V^0 candidates are identified through two tracks satisfying the following criteria:

- the significance of the distance of closest approach to the selected primary vertex in the transverse plane, $\mathcal{S}_d \equiv d/\sigma_d$ (see Sect. 5) of both tracks must satisfy $|\mathcal{S}_d| > 3$;
- the tracks' z coordinates at the point of closest approach in the transverse plane must be displaced from the primary vertex less than 1 cm, to suppress misreconstructed tracks;
- the resulting V^0 must have a distance of closest approach to the primary vertex of less than 200 μm. This requirement is intended to select only those V^0 candidates originating from the primary vertex, while candidates originating from heavy flavour decays may be taken into account during the tagging;
- the reconstructed mass should satisfy 472 MeV/$c^2 < m < 516$ MeV/c^2 for K_{S} candidates, and 1108 MeV/$c^2 < m < 1122$ MeV/c^2 for Λ candidates (in

the latter case, the higher p_{T} track is considered to be a proton; the other track, or both tracks in the case of K_{S} reconstruction, is assumed to be a charged pion). The mass peak plots of reconstructed K_{S} and Λ candidates are shown in Fig. 2.
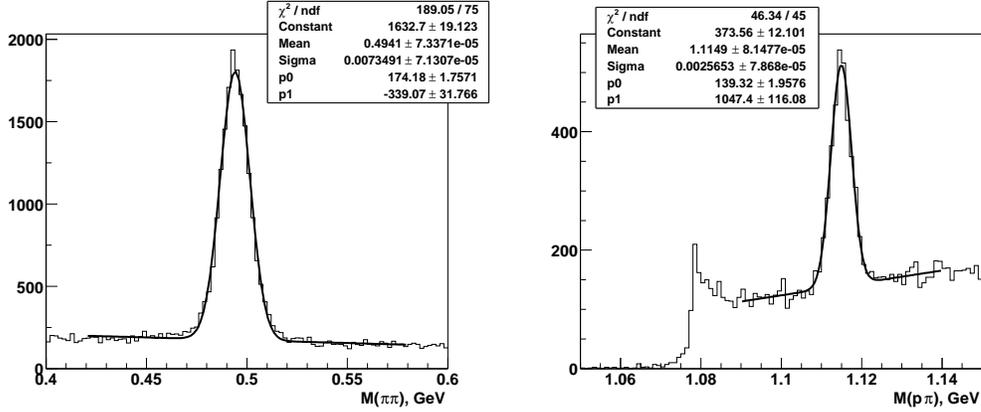
Fig. 2. Reconstructed K_{S} and Λ mass peaks.

Photon conversions are most easily recognized by the fact that the opening angle between the electron and positron is negligibly small. In the plane perpendicular to the beam line, this is exploited by requiring that the tracks should be less than 30 μm apart at the location where their trajectories are parallel to each other. In addition, they should again be oppositely charged, and their invariant mass is required to be less than 25 MeV/c^2. Since conversions happen inside material, the location of their vertices reflects the distribution of material inside the detector, as illustrated by Fig. 3.
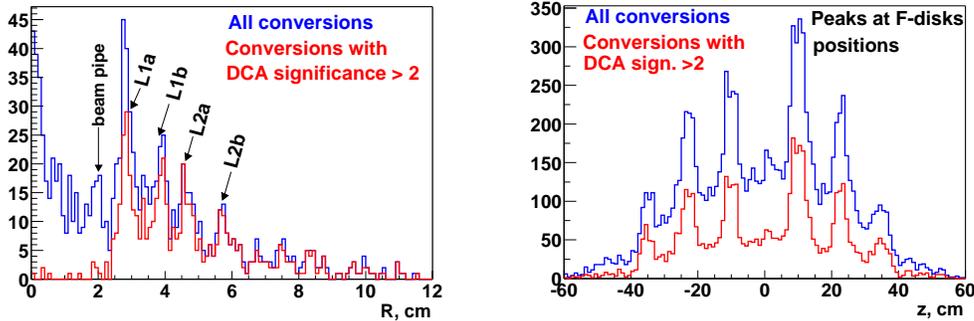
Fig. 3. Reconstructed radial (left) and z coordinate (right) of the conversion vertex.

It would be very desirable to know (roughly?) the V0 finding efficiency on the MC!

11

4 The Secondary Vertex Tag

The most intuitive tagging method is to require explicitly the presence of a displaced or *secondary* vertex. The requirement that a number of tracks can all be extrapolated to the same point in three dimensions is expected to lead to an algorithm that is robust even in the presence of misreconstructed tracks.

The reconstruction of secondary vertices starts from the track jet associated with each (taggable) calorimeter jet (see Sect. 3.2). It starts from all tracks from which that track jet was reconstructed, and first applies additional criteria to those tracks: they should have at least two SMT hits, transverse momenta exceeding 0.5 GeV/c, $|dca| < 0.15$cm, and $|zdca| < 0.4$cm. All tracks satisfying these criteria are used in a so-called *build-up* vertex finding algorithm. In detail, the algorithm consists of the following steps:

(1) vertices are reconstructed from all pairs of tracks using a Kalman vertex fitting technique [5], and are retained if the vertex fit yields a goodness-of-fit $\chi^2 < \chi^2_{max} = 100$;

(2) other tracks are added to the resulting vertex seeds one by one, and the combination yielding the smallest increase in fit χ^2 is retained;

(3) this procedure is repeated until the increase in fit χ^2 exceeds a set maximum, $\Delta\chi^2_{max} = 15$, or the total fit χ^2 exceeds χ^2_{max};

(4) the resulting vertex is selected if in addition, the angle ζ between the reconstructed momentum of the displaced vertex and the direction from the primary to the displaced vertex (in the transverse plane) satisfies $\cos\zeta > 0.9$; the vertex decay length in the transverse direction $L_{xy} < 2.6cm$.

(5) many displaced vertex candidates may result, with individual tracks possibly contributing to multiple candidates. The candidate list is pruned until no two candidates are associated with identical sets of tracks.

(6) Associate secondary vertices with calorimeter jets if $\Delta R(vtx, jet) < 0.5$.

In Fig. 4, we show distributions which characterize the properties of the secondary vertex reconstructed in $t\bar{t}$ MC events. We plot the multiplicity of vertices found in a track jet, the number of tracks associated with the vertex, the decay length significance of the vertex, and in case of multiple vertices in the jet, we also plot the largest decay length significance. The efficiency of this algorithm for a few different operating points as a function of the p_T and η of the associated calorimeter jet is shown in Fig. 5.

The algorithm as described above is referred to as the "Super Loose" algorithm in Sect. 7.1. It features a high efficiency but a relatively high fake rate. To exploit better the available information, a second tighter ("Loose") version of this algorithm is also applied. The differences with the first algorithm are:
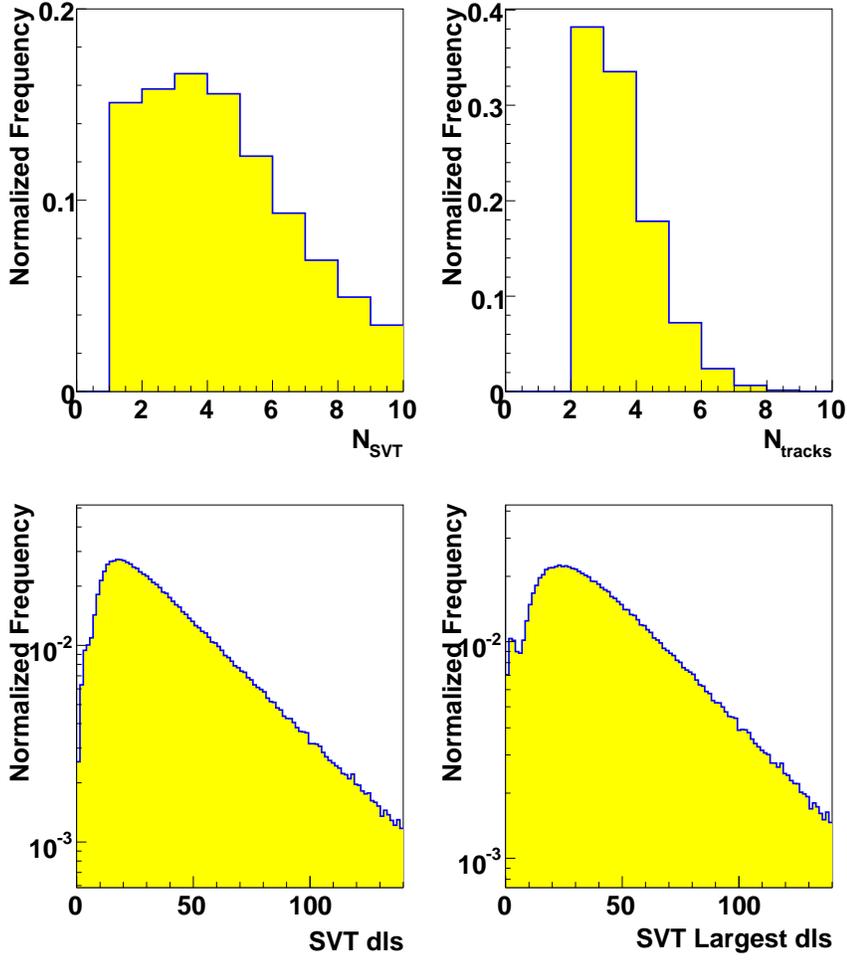
Fig. 4. Properties of the secondary vertices in $t\bar{t}$ MC: multiplicity of vertices found in a track jet (top left), the number of tracks associated with the vertex (top right), the decay length significance of the vertex (bottom left), and the largest decay length significance (bottom right).

- the input tracks are required to have a DCA significance $|\mathcal{S}_d| > 3$;
- the track association criterion uses $\Delta\chi^2_{\max} = 10$;
- the significance of the displacement of the vertex in the transverse plane must satisfy $\mathcal{S}_{xy} \equiv L_{xy}/\sigma(L_{xy}) > 5$.
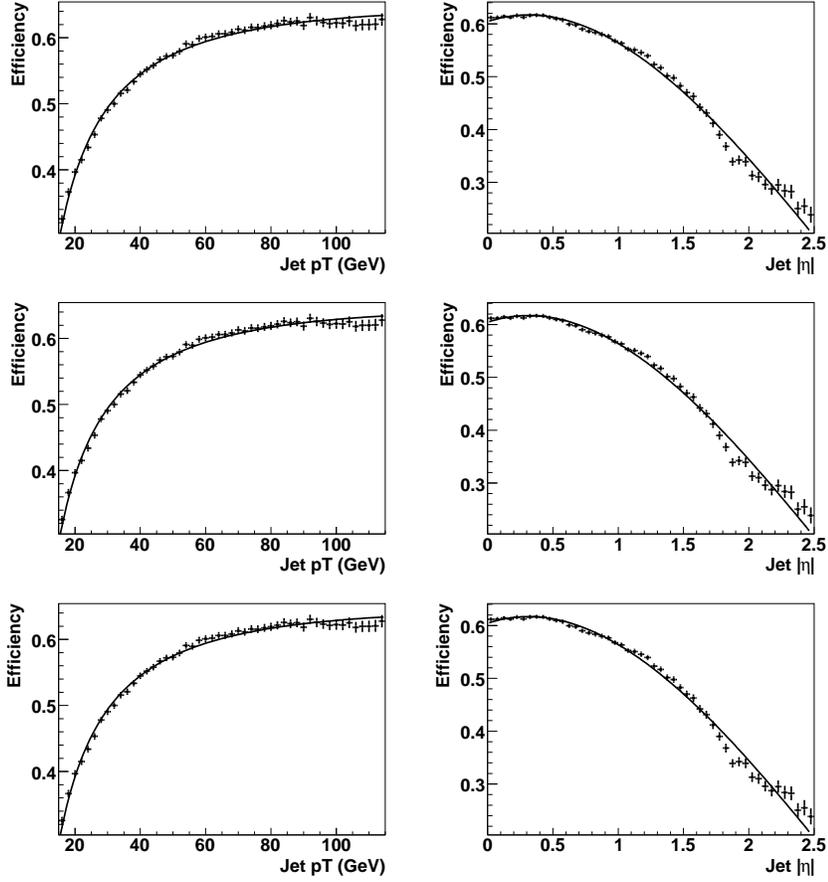
Fig. 5. Efficiency of the secondary vertex algroithm for tagging b-jets in $t\bar{t}$ MC events, as a function of the b-jet p_T and η for three different tagging criteria: "Loose" (top row), "Medium) (middle row) and "Tight" (bottom row).

5 The Jet LIfetime Probability (JLIP) tag

The impact parameters of all tracks associated to a calorimeter jet can be combined into a single variable, $\mathcal{P}_{\mathrm{JLIP}}$ [10], which can be interpreted as the probability that all tracks in a jet originate from the primary interaction point. Jets from light quark fragmentation are expected to present a uniform $\mathcal{P}_{\mathrm{JLIP}}$ distribution between 0 and 1, whereas jets from c and b quarks will exhibit a peak at very low $\mathcal{P}_{\mathrm{JLIP}}$ value. It is thus easy to select jets from b quarks by applying a cut on this probability, the level of the cut depending on the signal efficiency and background rejection desired for a given physics analysis.

Using the impact parameters of reconstructed tracks also allows the control of their resolution by using the real data themselves, minimizing the need for Monte-Carlo simulation. For this purpose, the impact parameter is signed by using the perigee coordinates of the track relative to the fitted primary vertex, $\vec{d}_{\mathrm{peri}}$, and the jet momentum vector, $\vec{p}_{\mathrm{T}}(\mathrm{jet})$. In the plane transverse to the beam axis, the distance of closest approach to the primary vertex ($d = |\vec{d}_{\mathrm{peri}}|$) is given the same sign as the scalar product $\vec{d}_{\mathrm{peri}} \cdot \vec{p}_{\mathrm{T}}(\mathrm{jet})$. The signed d distribution for tracks from light quark fragmentation is almost symmetric, whereas the distribution for tracks from b-hadron decay exhibits a long tail at positive values. Therefore, provided that the sign of $\vec{d}_{\mathrm{peri}} \cdot \vec{p}_{\mathrm{T}}(\mathrm{jet})$ is correctly determined, the negative part of the d distribution allows the d resolution function to be parametrized.

5.1 Calibration of the impact parameter resolution

In order to tune the computed error, σ_d^{raw}, on each track impact parameter, the following variable is introduced: $p_{\mathrm{scat}} = p(\sin\theta)^{3/2}$, where p is the particle momentum and θ its polar angle relative to the beam axis. In the plane transverse to the beam axis, the smearing due to multiple scattering is inversely proportional to $p_{\mathrm{T}} = p\sin\theta$ and proportional to the square root of the distance traveled by the track. Assuming the detector material to be distributed along cylinders aligned with the beam, this distance is also inversely proportional to $\sin\theta$. The d distributions are then computed in 16 different p_{scat} intervals.

In order to parametrize the d resolution, five track categories are considered:

- ≤ 6 CFT hits and ≥ 1 SMT hit (including the inner layer), for tracks with $|\eta| > 1.6$;
- ≥ 7 CFT hits and 1, 2, 3 or 4 SMT superlayer hits.

The first category includes forward tracks outside the CFT acceptance, the latter are central tracks with different numbers of SMT hits.
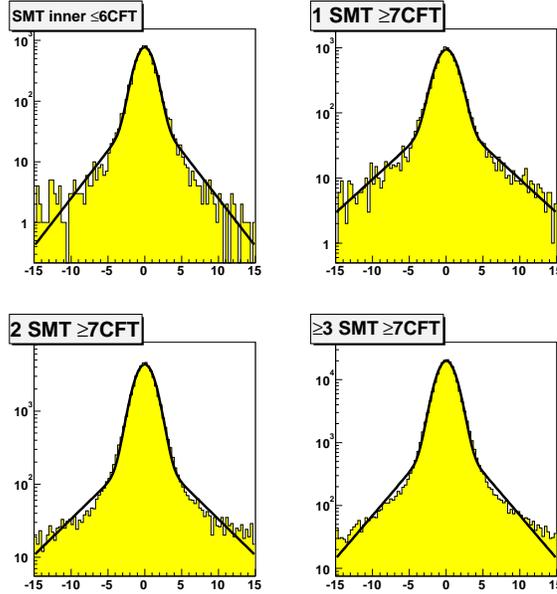
Fig. 6. Impact parameter significance, $\mathcal{S}^{\mathrm{raw}}$, distribution in the $p_{\mathrm{scat}} = p(\sin\theta)^{3/2}$ interval between 2.5 and 3 GeV/c, for different track categories in multi-jet data.

For each track, the significance is defined as the signed d value divided by its error:

$$\mathcal{S}_d^{\mathrm{raw}} = \frac{d}{\sigma_d^{\mathrm{raw}}} \ . \tag{1}$$

Figure 6 shows the $\mathcal{S}_d^{\mathrm{raw}}$ distribution for different track categories in a given p_{scat} interval. In each p_{scat} interval and each category, the $\mathcal{S}_d^{\mathrm{raw}}$ distribution is fitted using a Gaussian function (to describe the d resolution) and an exponential (to describe the tails). The fitted pull values (σ of the Gaussian in the previous fit) are presented in Figure 7 for multi-jet data and simulation. The superimposed curves are empirical parametrizations to the data and to the QCD Monte-Carlo. The pull values are found to go up to 1.2 in the data, while they are closer to 1 in the simulation.

As the impact parameter resolution may be sensitive to the primary vertex resolution, the d significance is also fitted separately for events with different numbers of tracks, N_{PV}, attached to the primary vertex. As shown in Figure 8 for multi-jet data, the pull value increases significantly with N_{PV} (here the p_{scat} dependence of the pull value is already corrected).

Then for each track, its d uncertainty can be corrected according to its measured p_{scat} value, category i and number of tracks N_{PV} at the primary vertex:

16
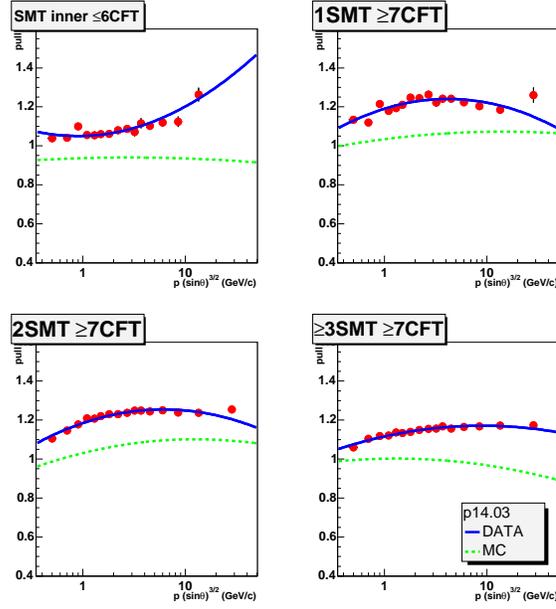
Fig. 7. Fitted pull values of the track impact parameters in multi-jet data (red points with error bars) as a function of $p_{\text{scat}} = p(\sin\theta)^{3/2}$, for different track categories. The superimposed blue solid (green dashed) curves are empirical parametrizations to the real data (QCD Monte-Carlo).
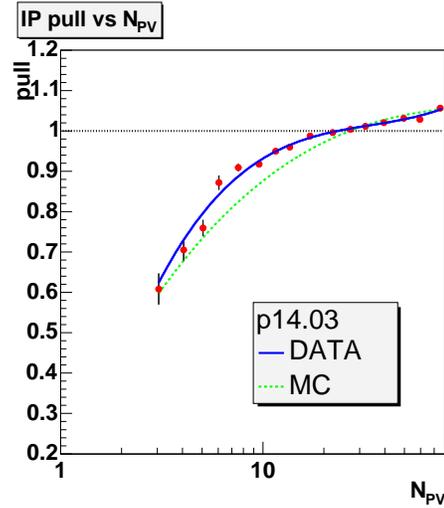
Fig. 8. Same as Figure 7 for the fitted pull values of the track impact parameters as a function of the number of tracks attached to the primary vertex, combining all track categories.

$$\sigma_d = pull(p_{\text{scat}}, i, N_{\text{PV}}) \cdot \sigma_d^{\text{raw}} \qquad (2)$$
$$\mathcal{S}_d = \mathcal{S}_d^{\text{raw}} / pull(p_{\text{scat}}, i, N_{\text{PV}}) \ .$$

17

The corrected σ_d resolutions are shown in Figure 9 for multi-jet data and for each category. They are parametrized as:

$$\sigma_d = \frac{a}{p \, (\sin \theta)^{3/2}} + b, \tag{3}$$

where a describes multiple scattering effects and b is the asymptotic resolution (which is sensitive to the primary vertex resolution, detector alignment, SMT intrinsic resolution, etc.). This parametrization is superimposed in Figure 9 for multi-jet data and for the QCD simulation.

For forward tracks with fewer than 7 CFT hits and with high p_{scat} ($> 10 \text{ GeV}/c$), the measured d resolution is in fact larger than its asymptotic fitted value (see Figure 9 upper left). **Do we know why that is??** More generally, the d resolution is better in the simulation than in the data.
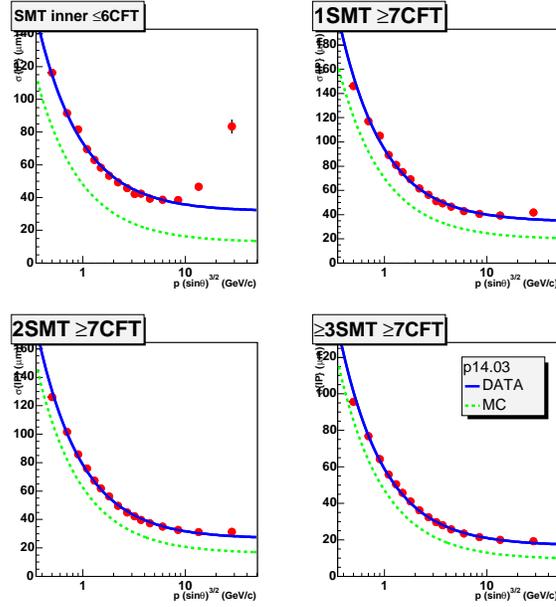
Fig. 9. Corrected impact parameter resolution (μm) for tracks in multi-jet data as a function of $p_{\text{scat}} = p(\sin \theta)^{3/2}$, for different track categories (points with error bars). The blue solid (green dashed) curve is a fit to the data (QCD Monte-Carlo).

5.2 Lifetime probability

The data themselves are used to calibrate the impact parameter significance. For multi-jet data or QCD Monte-Carlo, the negative part of the d significance distribution, denoted impact parameter resolution function $\mathcal{R}(\mathcal{S}_d)$, has been

parametrized as the sum of four Gaussian functions. The track categories used in the previous section are extended to take into account the number of SMT and CFT hits, $|\eta|$, χ^2 and p_T values of the tracks, as listed in Table 1. The category ranges are adjusted in order to describe as much as possible geometric and tracking effects, while keeping enough events in each category. For each of these 29 track categories, an impact parameter resolution function is fitted, as illustrated in Figure 10.
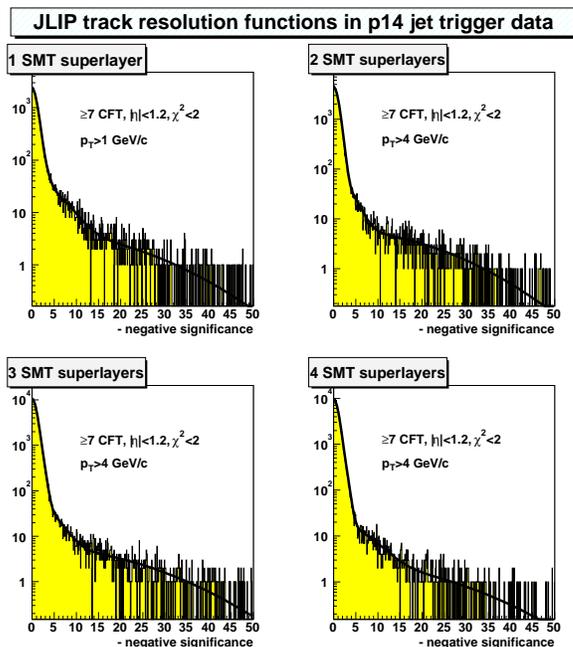
Fig. 10. Impact parameter resolution functions as measured in multi-jet data. They are shown here for four of the 29 track categories. The resolution functions are parametrized as the sum of 4 Gaussian functions.

SMT hits	CFT hits	$|\eta|$	χ^2	p_T (GeV/c)
≥ 1 hit in inner layer	≤ 6	1.6-2.0 or > 2.0	> 0	> 1
1 superlayer	≥ 7	< 1.2	0-2 or > 2	> 1
”	”	> 1.2	> 0	> 1
2, 3 or 4 superlayers	≥ 7	< 1.2	0-2	1-2, 2-4 or > 4
”	”	”	2-4 or > 4	> 1
”	”	1.2-1.6	0-2 or > 2	> 1
”	”	> 1.6	> 0	> 1

Table 1

Track categories used for the parametrization of the impact parameter resolution functions.

For tracks with a positive d significance, the resolution function can be con-

verted into a probability for this track to originate from the primary interaction point:

$$\mathcal{P}_{\text{trk}}(\mathcal{S}_d) = \frac{\int_{-50}^{-|\mathcal{S}_d|} \mathcal{R}(s)ds}{\int_{-50}^{0} \mathcal{R}(s)ds} \ , \tag{4}$$

where only tracks with $|\mathcal{S}_d| < 50$ are retained. **Here I'd like to say: "to reject further tracks from V^0 decays as well as misreconstructed tracks". Would that be correct?**

The corresponding track probabilities are shown in Figure 11 for multi-jet data and simulated jets of different flavours, and for positive and negative d values. Tracks with negative d values in multi-jet data and in simulated light quark jets are used to define the d resolution functions, thus ensuring uniform $\mathcal{P}_{\text{trk}}(\mathcal{S}_d < 0)$ probability distributions. For positive d, a significant peak at low $\mathcal{P}_{\text{trk}}(\mathcal{S}_d > 0)$ probability is present in simulated c- and b-jets. In multi-jet data, a peak is also observed at low values which is partly due to the presence of V^0's (which are not all removed), but also to tracks from charm and b-hadron decays recorded in these events. Note that for simulated c- and b-jets, a slight peak remains at negative d due to a flip of the d sign, mainly due to tracks very close to the jet axis direction.

Finally, the selected N_{trk} tracks with positive d significance are used to compute the *jet probability* $\mathcal{P}_{\text{JLIP}}$ as

$$\mathcal{P}_{\text{JLIP}} = \Pi \cdot \sum_{j=0}^{N_{\text{trk}}-1} \frac{(-\ln \Pi)^j}{j!} \quad \text{with} \quad \Pi = \prod_{i=1}^{N_{\text{trk}}} \mathcal{P}_{\text{trk}}(\mathcal{S}_{d,i}) \ . \tag{5}$$

For the tracks with negative d, a jet probability can be computed analogously (this will be relevant in Sect. 9).

By construction, if the $\mathcal{P}_{\text{trk}}$ are uniformly distributed and uncorrelated, the $\mathcal{P}_{\text{JLIP}}$ will also be uniformly distributed, independent of N_{trk}. Therefore, apart from wrongly assigned negative d in the case of tracks originating from the decay of long-lived particles, and from any correlations that are induced by the common primary vertex (which is reconstructed from the tracks under consideration, among others), the resulting $\mathcal{P}_{\text{JLIP}}$ distribution is indeed expected to be flat for negative d tracks in multi-jet data. These distributions are shown in Figure 12 for multi-jet data and simulated jets of different flavours, and for positive and negative d values.
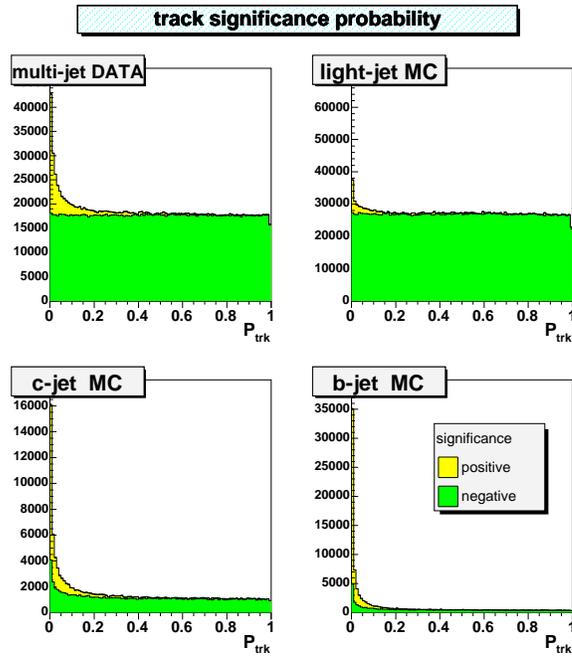
Fig. 11. Track probability in multi-jet data and qcd Monte-Carlo simulation of different flavours, for positive (light yellow) and negative (dark green) d values.
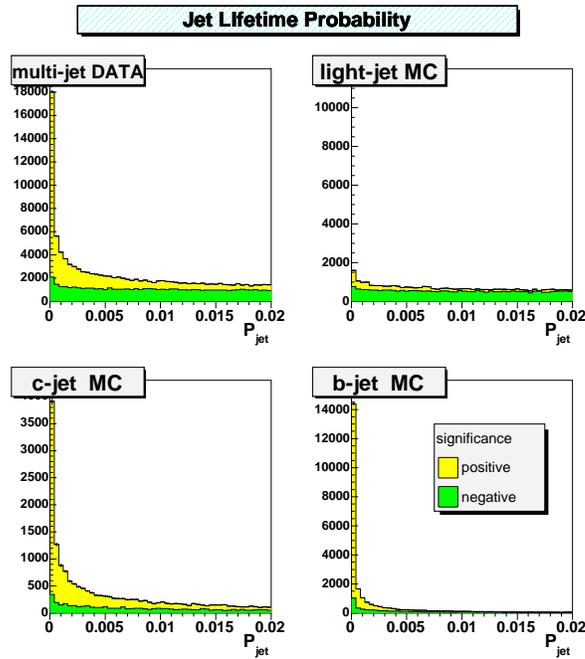
Fig. 12. Jet probability in multi-jet data and QCD Monte Carlo simulation of different flavours, for positive (light yellow) and negative (dark green) d values.

6 The Counting Signed Impact Parameter tag

In this method, like in Section 5, there is no attempt to use reconstructed secondary vertices. Instead, the signed impact parameter significance S_d is calculated for all good tracks located within a cone of radius 0.5 around the jet axis. For the present purpose, the definition of a good track is as follows:

- the track originates from the primary vertex (the difference between the z coordinates of the DCA point and the primary vertex is less than 1 cm);
- the track DCA must not be too large, $|d| < 2$ mm;
- the track transverse momentum should satisfy $p_T > 1$ GeV/c;
- the track fit should be of good quality, $\chi^2/N_{\mathrm{DF}} < 3$;
- at least 2 hits in the inner SMT layers or at least 3 hits anywhere in the SMT are required;
- tracks with fewer than 10 CFT hits (forward tracks) must have at least 4 stereo SMT hits.

These tracks are also required not to originate from a V^0 candidate, as detailed in Section 3.

A jet is considered to be tagged if there are at least two good tracks with $\mathcal{S}_d^{\mathrm{raw}}/a > 3$ or at least three good tracks with $\mathcal{S}_d^{\mathrm{raw}}/a > 2$, where a is a renormalizing parameter. The choice of a determines the operating point (b-tagging efficiency and mistag rate) of the algorithm. Alternatively, if a jet has at least two good tracks with positive $\mathcal{S}_d^{\mathrm{raw}}$, then the minimum value of a at which there are at least two good tracks with $\mathcal{S}_d^{\mathrm{raw}}/a > 3$ or at least three good tracks with $\mathcal{S}_d^{\mathrm{raw}}/a > 2$ can be used as a continuous output variable of the tagger. In the version of the algorithm used in DØ, a is fixed to be 1.2.

In the actual implementation of the algorithm, there is an additional condition related to the fact that the sign of $\mathcal{S}_d^{\mathrm{raw}}$ cannot be determined accurately for tracks that are very close to the jet axis. The criterion of closeness is empirically chosen as the difference in the azimuthal angle between the track and jet directions $\Delta\varphi$ being less than 20 mrad. Four categories of tracks are counted separately:

- tracks with $\mathcal{S}_d^{\mathrm{raw}}/a > 3$, $|\Delta\varphi| > 20$ mrad ("3σ-strong" tracks, their total number to be denoted as N_{3s}),
- tracks with $2 < \mathcal{S}_d^{\mathrm{raw}}/a < 3$, $|\Delta\varphi| > 20$ mrad ("2σ-strong" tracks, N_{2s}),
- tracks with $|\mathcal{S}_d^{\mathrm{raw}}/a| > 3$, $|\Delta\varphi| < 20$ mrad ("3σ-weak" tracks, N_{3w}),
- tracks with $2 < |\mathcal{S}_d^{\mathrm{raw}}/a| < 3$, $|\Delta\varphi| < 20$ mrad ("2σ-weak" tracks, N_{2w}).

If CSIP is used as a stand-alone algorithm, the jet is considered tagged if $N_{2s} + N_{3s} + N_{2w} + N_{3w} \geq 3$ and $N_{2s} + N_{3s} \geq 1$, or $N_{3s} + N_{3w} \geq 2$ and $N_{3s} \geq 1$. In the DØ implementation the four numbers are packed in a single variable

which is used in the combined algorithm as explained in Section 7.

7   The Neural Network tag

The Neural Network tag (or NN tag) combines input variables from the SVT, JLIP, and CSIP tagging algorithms to create a single b-tag. By using an artificial neural network, a more optimal combination of the variables is achieved. The NN implementation chosen is the TMULTILAYERPERCEPTRON implemented in the ROOT [11] framework.

The basic neural network structure is N_{input}:N_{hidden}:N_{output}. The number of hidden layers was chosen to be one, as it is advantageous to keep the NN as simple as possible. The NN has seven input nodes and one output node, and 24 hidden nodes in a single layer. The number of hidden nodes was optimized after choosing the input variables. **Can we substantiate the "advantage" of having only a single hidden layer?**

7.1   Input variables

The choice of input variables is crucial for the performance of the NN. Because of that, before creating the NN several studies were performed on the input variables. The initial variable sets, which were later re-optimized, were tested with a NN structure N:$2N$:1, where N is the number of input variables, using 500 training epochs and selection cuts of (SVT $\mathcal{S}_{xy} > 2$ or CSIP $\mathcal{N}_{\text{CSIP}} > 8$ or JLIP $\mathcal{P}_{\text{JLIP}} < 0.02$). The $\mathcal{N}_{\text{CSIP}}$ variable is defined below.

Seven input variables were selected for their good discrimination between b-jets and non-b jets. Five of the variables are based on the secondary vertices reconstructed using the SVT algorithm. The remaining two summarize information from the JLIP and CSIP algorithms. In detail, the input variables are:

SVT $\mathcal{S}_{xy}$ - The decay length significance (the decay length in the transverse plane divided by its uncertainty) of the secondary vertex with respect to the primary vertex.

SVT χ^2_{dof} - The χ^2 per degree of freedom of the secondary vertex fit.

SVT N_{trk} - The number of tracks used to reconstruct the secondary vertex.

SVT m_{vtx} - The mass of the secondary vertex. The mass is calculated from the combined momentum four-vectors of the tracks, assuming all particles were pions.

SVT N_{vtx} - The number of secondary vertices reconstructed in the jet.

JLIP $\mathcal{P}_{\text{JLIP}}$ - The "jet lifetime probability" computed in Section 5.

CSIP $\mathcal{N}_{\text{CSIP}}$ - A combined variable based on the number of tracks with an impact parameter significance greater than some value. This variable is discussed in more detail below.

24

Since more than one secondary vertex can be found for each jet, vertex variables are ranked in order of the most powerful discriminator, the decay length significance ($\mathcal{S}_{xy}$). The secondary vertex with the largest $\mathcal{S}_{xy}$ in a jet is used to provide the input variables. If no secondary vertex is found, the SVT values are set to 0, apart from the SVT χ^2_{dof} which is set to 75 corresponding to the upper bound of χ^2_{dof} values.

A rather standard implementation of a secondary vertex tagging algorithm would require a significantly displaced vertex, possibly along with other quality criteria. While such an approach helps to isolate a pure sample of heavy flavour decays, it typically results in a low efficiency. In the context of a NN optimization, this is undesirable as any vertex-related information is only available if a displaced vertex is found. For this reason, the "Super Loose" SVT algorithm as described in Sect. 4 is used: even if the vertex candidates it finds are of significantly poorer quality, it finds many candidates **(it would be good to be able to refer to a figure here!)**, and they still provide additional discrimination between b-jets and other flavours. Figure 13 shows the efficiency for b-quark and light-flavour jets of both algorithm choices on QCD samples. The NN tagger is found to perform best if information for both the "Super Loose" and a tighter SVT algorithm, the "Loose" version also discussed in Sect. 4, is used: the N_{trk} variable is taken from the latter, and all other SVT variables from the former.
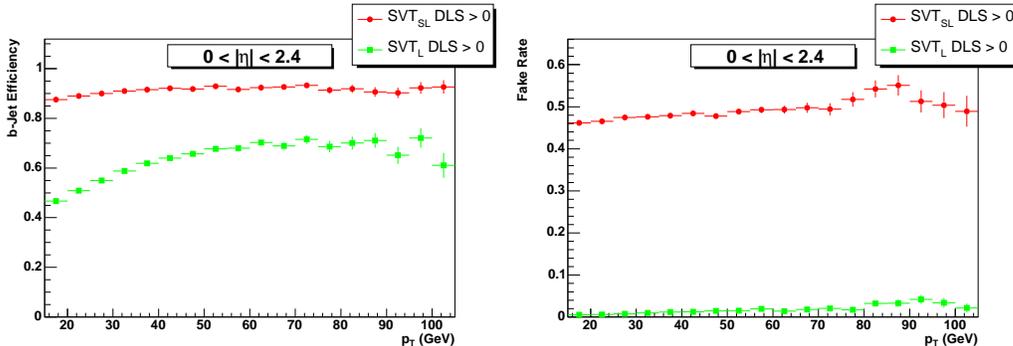
Fig. 13. SVT_{SL} (red circles) and SVT_L (green squares) efficiencies for QCD $b\bar{b}$ and QCD fake MC jets.

The CSIP $\mathcal{N}_{\mathrm{CSIP}}$ variable is based on the four CSIP variables N_{3s}, N_{2s}, N_{3w}, N_{2w} described in Section 6. Neural Networks tend to perform best when provided with continuous values spread over a range. Since the CSIP variables have small integer values which are not very good as inputs, they are combined in one variable which brings the advantage of reducing the number of input variables, hence simplifying the NN:

$$\mathcal{N}_{\mathrm{CSIP}} = 6 \times N_{3s} + 4 \times N_{2s} + 3 \times N_{3w} + 2 \times N_{2w}. \tag{6}$$

The weights were determined in an empirical manner to give optimum perfor-

mance for this variable alone.

The distributions for the input variables to the NN in Monte Carlo and QCD data are shown in Fig. 14.
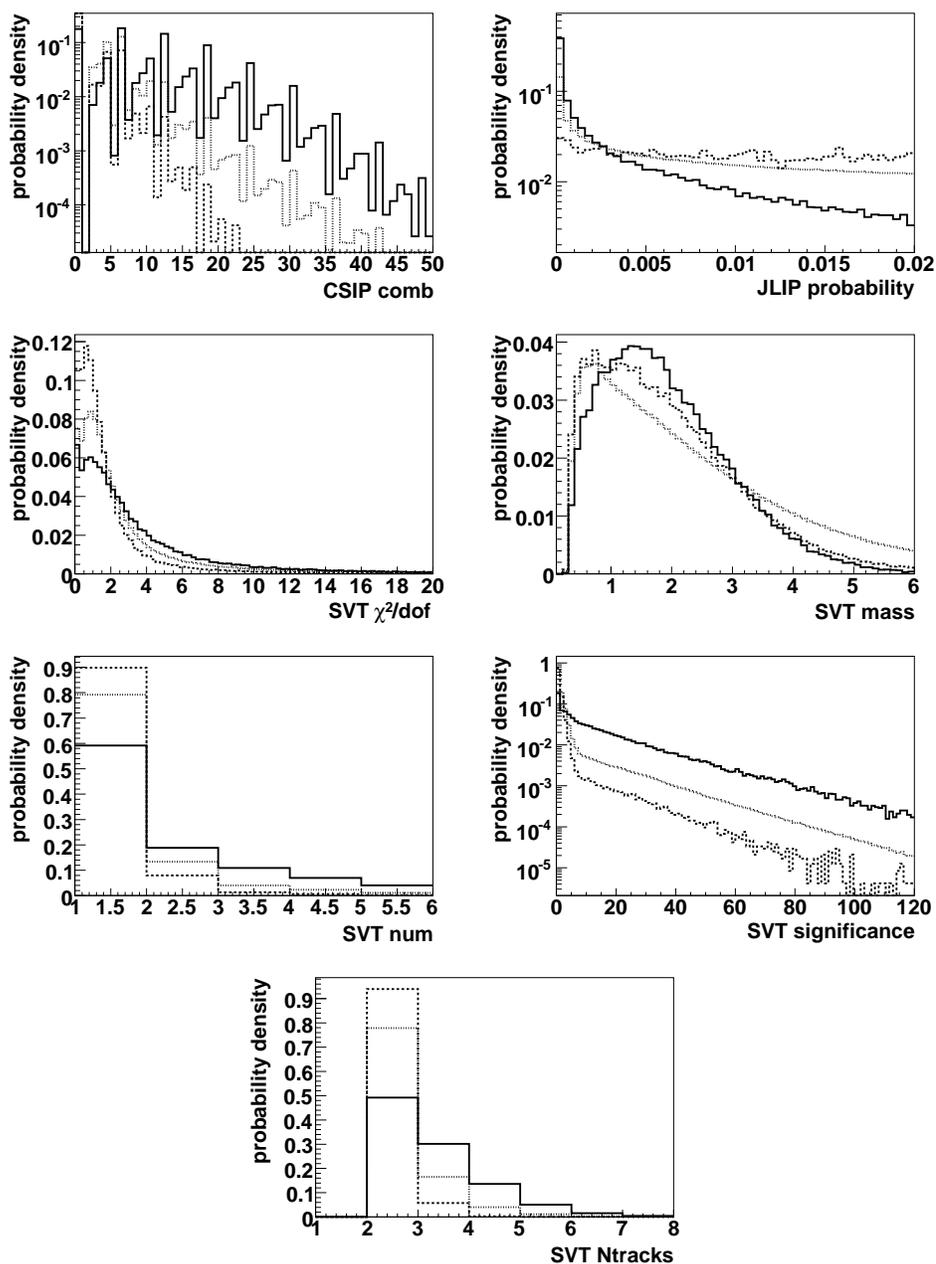
Fig. 14. The NN variables CSIP $\mathcal{N}_{\mathrm{CSIP}}$, JLIP $\mathcal{P}_{\mathrm{JLIP}}$, SVT χ^2_{dof}, m_{vtx}, N_{vtx}, S_{xy} and N_{trk} for QCD $b\bar{b}$ MC (solid lines), light jet QCD MC (dashed lines) and data (dotted lines). All histograms are normalized to unity.

7.2   Neural Network Training

The training samples and training algorithms are optimized prior to the input variable optimization as they do not depend on the input variable optimization. The NN is trained on QCD light jet and $b\bar{b}$ samples, weighted to give an equal number of jets after input selection cuts. The signal sample of 270,000 $b\bar{b}$ events and the background sample of 470,000 light jet QCD events are each split in half, with one half used for the training sample and the other half for the test sample.

The number of training epochs was varied from 50 up to 2000.After testing different training versions, the number of training epochs was set to 400.

7.3   Input Selection Cuts

Another important attribute of the NN is the selection of the jets which are used to train the NN. A selection which is too loose can cause a loss of performance and resolution as the NN is learning a signal from background separation which could have been carried out with a simple cut. A selection which is too tight will cause a significant loss of b-jets and therefore limit the NN training.

The input selection cuts were optimized by considering each variable in turn, optimizing first SVT $\mathcal{S}_{xy}$, then JLIP $\mathcal{P}_{\mathrm{JLIP}}$, and finally CSIP $\mathcal{N}_{\mathrm{CSIP}}$. The NN was trained on the QCD $b\bar{b}$ and QCD light-flavour samples and the optimization plots were produced from a high p_T `ALPGEN` $t\bar{t}$ sample and cross checked with the QCD $b\bar{b}$ sample to ensure there was no sample, p_{T} or MC generator dependence in the optimization.

7.4   NN Performance on Monte Carlo

The output from the optimized NN b-tagger on $b\bar{b}$ and light-flavour QCD MC is shown in Fig 7.4. The NN b-tagger shows significant separation between signal and background samples.

7.5   Operating Points

The NN tagger performance was evaluated for 12 operating points, as detailed in Table 2.
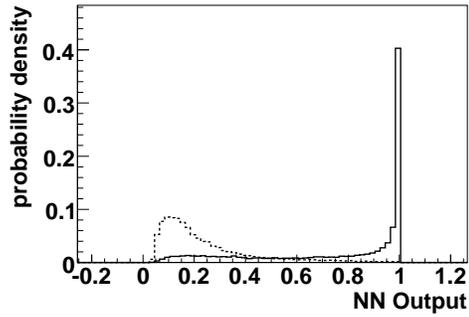
Fig. 15. The NN output for the light jet QCD MC (dashed line) compared to the QCD $b\bar{b}$ MC (solid line). Both distributions are normalized to unity.

Table 2
The NN tagger's operating points.

Name	MegaTight	UltraTight	VeryTight	Tight	Medium	oldLoose
NN Cut	> 0.925	> 0.9	> 0.85	> 0.775	> 0.65	> 0.5

Name	Loose	L2	L3	L4	L5	L6
NN Cut	> 0.45	> 0.325	> 0.25	> 0.2	> 0.15	> 0.1

8   Efficiency estimation

The performance of the tagging algorithm cannot simply be inferred from simulated samples. Several effects cause differences between simulated and real data:

- Monte Carlo hit resolutions, both in the CFT and in the SMT, have been tuned to reproduce those in the data. However, the tuning cannot be expected to be perfect as the observed resolutions in the data are also affected by un-understood geometrical effects which do not play a role in the simulation;
- a small but non-negligible fraction of the channels, in particular in the SMT, fail to readout from time to time.

These effects lead to different effective resolutions and efficiencies. A calibration is therefore required. To this end, b- and c-jets are denoted "signal" and their efficiency estimation is described below. Jets originating from u, d, or s quarks or gluons are considered background, and their tag rate estimation is described in Section 9.

8.1   The SystemD method

The *SystemD* method has been developed to determine identification efficiencies using almost exclusively the real data. Monte Carlo simulation is only used to estimate correlation coefficients. The method involves several, essentially uncorrelated, identification criteria which are applied to the same data sample. Combining these criteria allows the definition of a system of equations which can be solved to extract the efficiency of each criterion.

The data sample is assumed to be composed of a signal and several backgrounds. Let f_0 be the fraction of signal events and $f_{i=1..b}$ the fraction of each considered background where b is the total number of backgrounds. These fraction must satisfy:

$$\sum_{i=0}^{b} f_i = 1. \tag{7}$$

Then consider c uncorrelated identification criteria with different selection efficiencies $\varepsilon_{i=0..b}^{k=1..c}$ on the signal and backgrounds. Only a fraction Q^k of the total number of events will pass the k-th identification criterion. Then a new

29

set of equations can be added for each selection:

$$\sum_{i=0}^{b} \varepsilon_i^k f_i = Q^k. \tag{8}$$

As the selection criteria are uncorrelated, the total efficiency $\varepsilon_{i=0..b}^{k_1,..,k_r}$ $(r \leq c)$ of applying successively several of them can be factorized in terms of single efficiencies ε_i^k :

$$\varepsilon_i^{k_1,..,k_r} = \prod_{v=1}^{r} \varepsilon_i^{k_v}. \tag{9}$$

A generalization of Eqn. (8) can then be obtained for a combination of several criteria:

$$Q^{k_1,..,k_r} = \sum_{i=0}^{b} \Big(\prod_{v=1}^{r} \varepsilon_i^{k_v} \Big) f_i. \tag{10}$$

The signal and background fractions are $b+1$ unknown parameters and each identification criterion introduces $b+1$ new unknowns in the form of selection efficiencies. The number of equations of the form (10) depends on the number of combinations of the c criteria which leads to a total of $\sum_{r=0}^{c} \binom{r}{c} = 2^c$ equations. To obtain a system of equations which can be solved, b and c must satisfy:

$$(1+c) \times (1+b) \leq 2^c. \tag{11}$$

The simplest non-trivial solutions are:

- $c = 3$, $b = 1$: 8 equations with 8 unknowns;
- $c = 4$, $b = 2$: 16 equations with 15 unknowns.

The system of equations is nonlinear and can have several solutions. Only the simplest case of 8 equations will be considered in the following. This system has two solutions, which differ by the interchange of efficiencies assigned to the signal and background samples. As will be detailed in Sect. 8.2, further *a priori* knowledge of at least one the unknown parameters is required to resolve the ambiguity. The input parameters are the fractions of events $Q_i^{k_0,..,k_r}$ which are determined directly from the real data. There is therefore no input from simulated events. Solving the system gives access to the signal and background fractions and to the various efficiencies.

In a more realistic model, identification criteria can have some correlation between them. Such correlations can be accounted for by adding into the

equations correction factors, which then need to be evaluated independently. In the application described in Sect. 8.2, they will be estimated on simulated data sets. Generically, a correction factor κ_i^{jk} describing the degree of correlation between the selections j and k applied to sample i is obtained as

$$\kappa_i^{jk} = \frac{\varepsilon_i^{jk}}{\varepsilon_i^j \varepsilon_i^k} = \frac{N_i \, N_i^{jk}}{N_i^j N_i^k}, \tag{12}$$

where N_i represents the total number of (simulated) events for sample i, N_i^j and N_i^k denote the numbers of events passing selections j and k, and N_i^{jk} denotes the number of events pass both selections. Higher order correlations can be introduced in a similar fashion. Summarizing, the method remains mostly based on real data. The Monte Carlo simulation is only used to estimate correction factors which are ratios of efficiencies.

8.2 Application to b-tagging efficiency measurements

The *SystemD* methods is used here in order to extract the b-tagging efficiencies of the NN tagger. The method is applied to a sample of jets in real data events satisfying the following criteria:

- $E_{\mathrm{T}} > 15$ GeV;
- $|\eta| < 2.5$;
- the jet must be taggable;
- the jet must contain a muon with $p_T^\mu > 4$ GeV/c within a cone of radius $R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.5$ around the jet axis. **The original (JLIP) text stated a cut $R < 0.7$?? And how about the fact that we require Medium muons with $\chi^2 < 100$ here – go into detail?**

The lifetime composition of the resulting sample could be biased by third level trigger requirements applying impact parameter or secondary vertex requirements. To avoid such biases, events are required to have passed at least one lifetime-unbiased trigger. These requirements result in a sample of $141 \cdot 10^6$ jets **(does this include the requirement of an away jet??)**. The sample consists of a mixture of b-, c-, and light-flavour jets. The first two are mostly due to semimuonic decays of B and C hadrons; muons in light-flavour jets arise mainly from in-flight decays of $\pi^\pm$ and K$^\pm$ mesons. To apply the *SystemD* method as described above, however, only a single source of background can be dealt with. The c- and light-flavour backgrounds are therefore lumped together in the following. An important consequence of this is the fact that the use of the *SystemD* method only allows to determine the efficiency for a specific mixture of c-quark and light-flavour jets; it is therefore not useful to extract efficiencies for the separate background sources.

Three identification criteria are used which include the working point of the b-tagging algorithm under study. The two other selections are based on other properties of b-jets events.

The first one is a cut on the transverse momentum of the muon relative to the direction obtained by adding the muon and jet momenta, $p_{\mathrm{T,rel}}$. This criterion is chosen because the high $p_{\mathrm{T,rel}}$ values in B hadron decays are due to the high mass of the b quark, and as such are in principle expected to be independent of the lifetime criterion; the corresponding correction factors (see Sect. 8.1), denoted κ_{b} for b-jets and κ_{cl} for background jets, are evaluated below. A cut $p_{\mathrm{T,rel}} > 0.5$ GeV/c is chosen in *SystemD* in order to have a similar efficiency for c and light quark jets and to have a limited sensitivity to the $p_{\mathrm{T,rel}}$ distribution tail. Any sensitivity to this value is taken into account as a systematic uncertainty.

As b quarks are usually produced in pairs, the last selection criterion consists of increasing the fraction of b-jets by looking for another tagged jet in the same event, using the JLIP algorithm (see Section 5) and requiring that $\mathcal{P}_{\mathrm{JLIP}} < 0.005$ for this jet. As this criterion is not applied to the muon jet itself, no correlation with the $p_{\mathrm{T,rel}}$ criterion is expected; this hypothesis has been verified explicitly. The lifetime tagging requirements applied to both jets, however, could be correlated by the fact that they involve the same primary vertex. The corresponding correction factors are denoted β for b-jets and α for background jets, and are again evaluated below. It should be pointed out, however, that the application of the $\mathcal{P}_{\mathrm{JLIP}}$ cut modifies the flavour composition of the background sample, as the charm tagging efficiency is expected to be significantly higher than that for light-flavour jets. This causes a dependence of α on the physics assumptions made in the Monte Carlo programs (the production cross sections for the various processes). Fortunately, it turns out that the uncertainty on α affects the b-tagging efficiency only very marginally.

To illustrate the effect of these additional cuts, the $p_{\mathrm{T,rel}}$ distribution is shown in Figure 16: once for all taggable jets (top), and once after applying the $\mathcal{P}_{\mathrm{JLIP}}$ cut both to the muon jet and to another taggable jet observed in the same event (bottom). The data are fitted to Monte-Carlo templates of each quark flavour, with a free normalization. A reasonable agreement is obtained; a clear increase in the fraction of b jets is observed upon applying the $\mathcal{P}_{\mathrm{JLIP}}$ cut. (It should be pointed out that identical templates are used in both cases: although such fitting procedure could in principle also be used to estimate b-jet tagging efficiencies, it makes more assumptions than the *SystemD* method and is not used for this purpose.) **Here I think it is inappropriate to show a plot where the away-side cut is applied to both jets: that does not correspond to the actual procedure!**

32
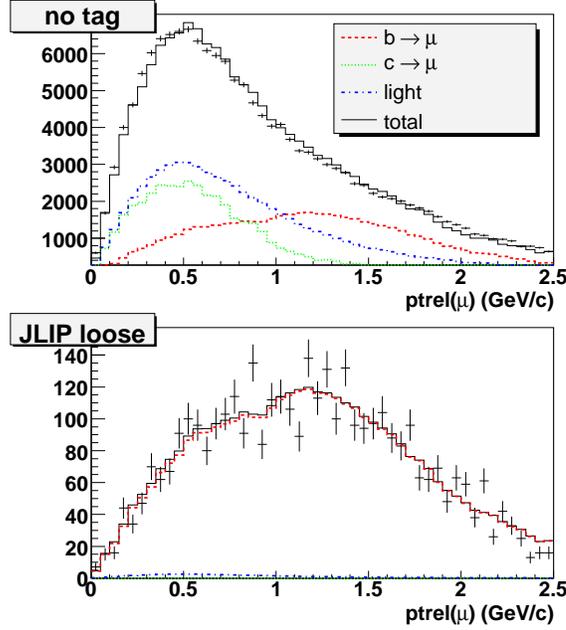
Fig. 16. $p_{\mathrm{T,rel}}$ distributions of muons in taggable jets: (top) without lifetime tagging, (bottom) with a $\mathcal{P}_{\mathrm{JLIP}}$ cut applied both to the muon jet and to another taggable jet observed in the same event. The superimposed fitted histograms take into account the contribution of $b \to \mu X$ (red dashed), $c \to \mu X$ (green dotted), light quark jets (blue dash-dotted), and their sum (full histogram).

Denoting the criteria used in *SystemD* as (t) for the lifetime tagging criterion, (m) for the $p_{\mathrm{T,rel}}$ cut, and (b) for the $\mathcal{P}_{\mathrm{JLIP}} < 0.005$ cut applied to another taggable jet, and with the notation for the correction factors as above, the final system to solve is therefore:

$$
\begin{aligned}
f_{\mathrm{b}} & + f_{\mathrm{cl}} & = 1 \\
f_{\mathrm{b}}\varepsilon_{\mathrm{b}}^{t} & + f_{\mathrm{cl}}\varepsilon_{\mathrm{cl}}^{t} & = Q^{t} \\
f_{\mathrm{b}}\varepsilon_{\mathrm{b}}^{m} & + f_{\mathrm{cl}}\varepsilon_{\mathrm{cl}}^{m} & = Q^{m} \\
f_{\mathrm{b}}\varepsilon_{\mathrm{b}}^{b} & + f_{\mathrm{cl}}\varepsilon_{\mathrm{cl}}^{b} & = Q^{b} \\
f_{\mathrm{b}}\kappa_{\mathrm{b}}\varepsilon_{\mathrm{b}}^{t}\varepsilon_{\mathrm{b}}^{m} & + f_{\mathrm{cl}}\kappa_{\mathrm{cl}}\varepsilon_{\mathrm{cl}}^{t}\varepsilon_{\mathrm{cl}}^{m} & = Q^{t,m} \\
f_{\mathrm{b}}\varepsilon_{\mathrm{b}}^{m}\varepsilon_{\mathrm{b}}^{b} & + f_{\mathrm{cl}}\varepsilon_{\mathrm{cl}}^{m}\varepsilon_{\mathrm{cl}}^{b} & = Q^{m,b} \\
f_{\mathrm{b}}\beta\varepsilon_{\mathrm{b}}^{b}\varepsilon_{\mathrm{b}}^{t} & + f_{\mathrm{cl}}\alpha\varepsilon_{\mathrm{cl}}^{b}\varepsilon_{\mathrm{cl}}^{t} & = Q^{b,t} \\
f_{\mathrm{b}}\kappa_{\mathrm{b}}\beta\varepsilon_{\mathrm{b}}^{t}\varepsilon_{\mathrm{b}}^{m}\varepsilon_{\mathrm{b}}^{b} & + f_{\mathrm{cl}}\kappa_{\mathrm{cl}}\alpha\varepsilon_{\mathrm{cl}}^{t}\varepsilon_{\mathrm{cl}}^{m}\varepsilon_{\mathrm{cl}}^{b} & = Q^{t,m,b}
\end{aligned}
\tag{13}
$$

At this point, a remark is in order: as the *SystemD* method leads to a set of *nonlinear* equations, two possible solutions exist for the quantity of interest, the b-tagging efficiency. The ambiguity between these two solutions uses the a

33

priori knowledge that the efficiency of each selection criterion for b-jets should be higher than for background jets: $\varepsilon_b^{t,m,b} > \varepsilon_{cl}^{t,m,b}$.

8.3 SystemD Correction Factors

I'm not sure this figure is still useful, given that we actually do take α from the MC!
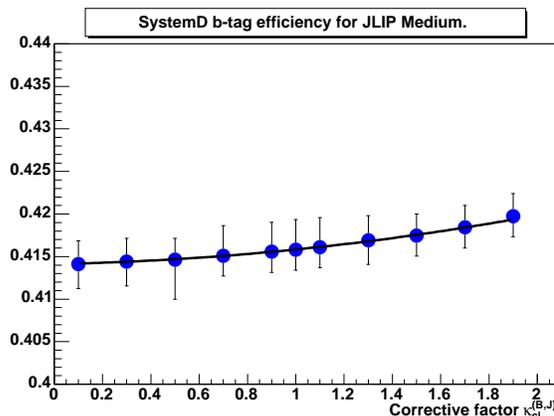
Fig. 17. Efficiency of the JLIP Medium tagging algorithm measured with the *SystemD* method as a function of the α correction factor. The other input parameters are given in Table 3.

Table 3 gives an example of *SystemD* applied to the JLIP Medium working point on a sample of roughly 8 millions of selected jets: fractions of jets passing each identification criterion, correlation coefficients and solutions of the system (13).

We need plots for the MC-derived correlation coefficients here!

8.4 Further Corrections

The b-tagging efficiency obtained with the *SystemD* method is valid for jets with a semi-muonic decay of the b quark. To obtain the efficiency for an inclusive sample, a correction is determined from simulated $Z \to b\bar{b}$ events with B hadrons decaying inclusively or as $b \to \mu X$. The final efficiency is then defined as:

$$\varepsilon_b^{\text{data}} = \frac{\varepsilon_b^t \cdot \varepsilon_b^{\text{MC}}}{\varepsilon_{b \to \mu X}^{\text{MC}}} = SF_b \cdot \varepsilon_b^{\text{MC}} , \qquad (14)$$

Table 3
Numerical example of *SystemD* applied to the JLIP Medium tagger. The first column gives the fractions of selected jets which are the system input. The solutions of the system is presented in the eight upper right array elements (bold). The measured b-tagging efficiency for JLIP Medium is 42%. The eight lower right elements present the correlation coefficients estimated from the simulation. The factors κ_b, κ_{cl} and α are fixed to 1. The correlation coefficients $\kappa_b^{t,m,b}$ and $\kappa_{cl}^{t,m,b}$ are approximated as the product of the two-criteria correlation coefficients. **(Drop this last sentence??)**

Total			
	Q^0 fraction	f_b fraction	f_{cl} fraction
	1	**0.26**	**0.74**
Single selection criteria			
(x)	$Q^{(x)}$ fraction	ε_b efficiency	ε_{cl} efficiency
T	0.13	**0.42**	**0.025**
M	0.45	**0.66**	**0.37**
B	0.062	**0.14**	**0.033**
Combined selection criteria			
(x)	$Q^{(x)}$ fraction	κ_b factor	κ_{cl} factor
T, M	0.077	0.99	0.90
M, B	0.034	1	1
B, T	0.016	1.016	1
T, M, B	0.011	1.006	0.90

where $SF_b = \varepsilon_{b\to\mu X}^{data}/\varepsilon_{b\to\mu X}^{MC} = \varepsilon_b^t/\varepsilon_{b\to\mu X}^{MC}$ is the data-to-simulation efficiency scale factor. The tagging efficiency for c quark jets is not measured in data. It is assumed that the data-to-simulation scale factor is identical for b and c-jets. The c-jet tagging efficiency is then derived from the simulation by:

$$\varepsilon_c^{data} = SF_b \cdot \varepsilon_c^{MC}. \tag{15}$$

8.5 SystemD cross-check

Is it useful to retain this section? If so, it should be redone with the p17 NN tagger!

The validity of the *SystemD* method has been checked on a sample of simulated events. In such a sample the jet flavour is known and the *SystemD* result can

be compared to the Monte-Carlo truth. As too few light jets contain a muon, a random track within the jet cone with $p_T > 4$ GeV/c is used to model a muon from pion or kaon decay. The test sample is composed of 430 000 jets with a muon (75% light-flavour jets, 5% c-jets, and 20% b-jets). *SystemD* has been applied in seven bins of jet transverse momentum, from 15 to 135 GeV/c. The JLIP Medium working point is used.Its correlation coefficients are given in Table 3. The results are presented in Figure 18 and shows a good agreement with the expected values, especially regarding the JLIP b-tagging efficiency.
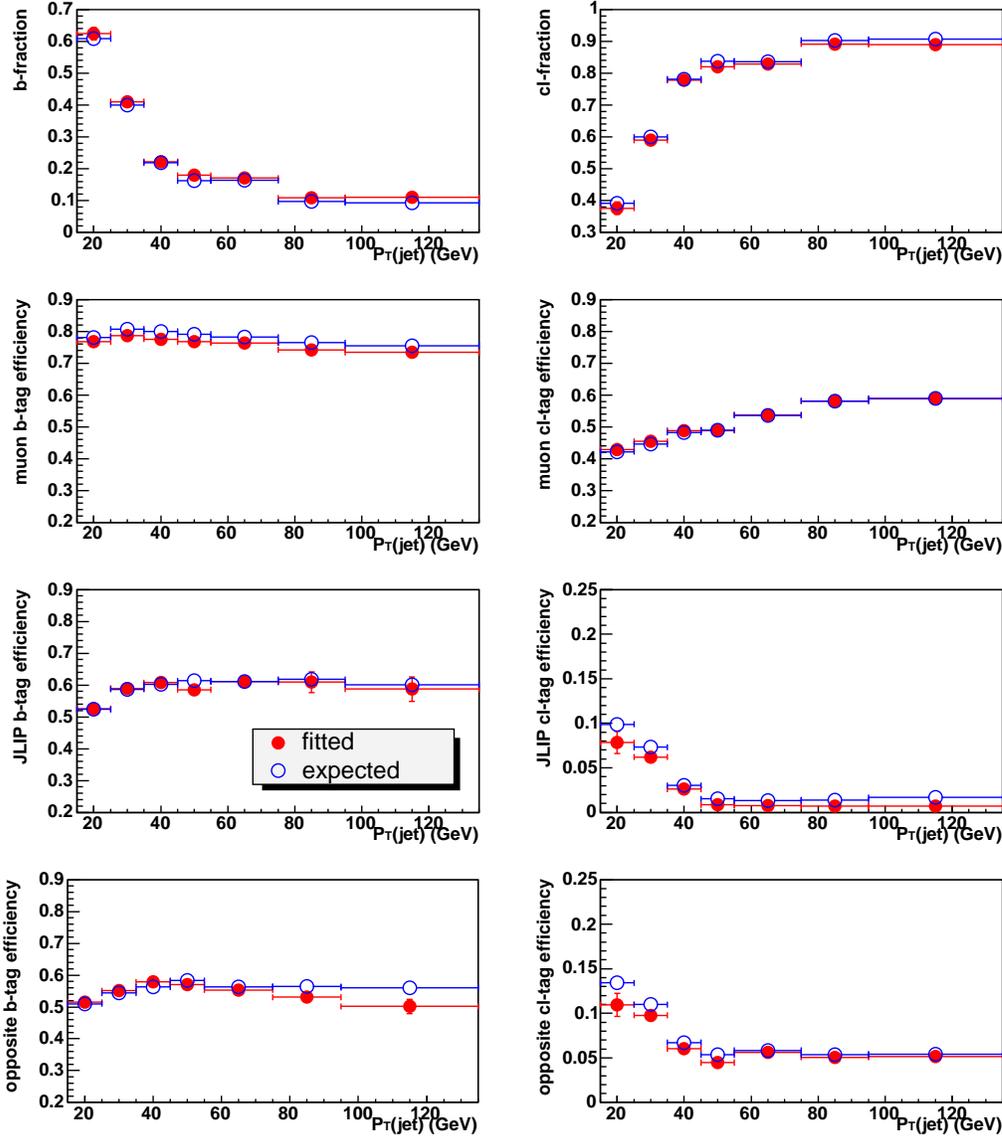
Fig. 18. Comparison between the *SystemD* results (full dots) and the Monte-Carlo truth (open dots) for a test sample of simulated jets for JLIP Medium as a function of the jet transverse momentum. From left to right, top to bottom: f_b, f_{cl}, ε_b^m, ε_{cl}^m, ε_b^t, ε_{cl}^t, ε_b^b, ε_{cl}^b.

8.6   Data and MC Samples

The MC samples used are outlined in Table 4. **Removed the add'l selection criteria here – already mentioned. Move (or remove) the MC samples table, too?**

Table 4
The MC samples used in certification of the NN tagger.

Sample	Number of Events
inclusive $t\bar{t}$	571,750
$Z \to b\bar{b}$	105,250
$Z \to b\bar{b}$ with μ	105,750
$Z \to c\bar{c}$	107,250
$Z \to c\bar{c}$ with μ	107,750
$Z \to q\bar{q}$ (q = u,d,s)	103,750
$Z \to q\bar{q}$ with μ (q = u,d,s)	107,000
QCD $p_T = 20 - 40, 40 - 80, 80 - 160, 160 - 320$	972,500
QCD $b\bar{b}$ $p_T = 20 - 40, 40 - 80, 80 - 160, 160 - 320$	265,000
QCD $c\bar{c}$ $p_T = 20 - 40, 40 - 80, 80 - 160, 160 - 320$	239,287

8.7   Jet Samples

The number of jets available in the MC samples after data processing and physics object and jet cut selections is outlined in Table 5. Only taggable jets (see Section 3.2) are used, and only events with at least two taggable jets are considered.

The different energy QCD samples are merged into a continuous p_T "QCD Merged" sample by weighting the different energy samples so that a continuous fall-off in the jet p_T spectrum is obtained. Larger b and c-jets samples are created by combining all the appropriate jets from the different production channels.

The precise relevance of this subsection isn't entirely clear to me. Would it be more appropriate in the NN section?

Table 5

Number of MC jets of each flavour available in each of the samples after data processing and jet selection.

Flavour	Sample	Number of Jets
b	$Z \to b\bar{b}$	130,220
	$t\bar{t} \to b$	1,662,929
	QCD $b\bar{b}$ (Merged)	90,252
	b (Combined)	2,349,755
$b \to \mu X$	$Z \to b\bar{b} \to \mu X$	42,726
	$t\bar{t} \to b \to \mu X$	175,579
	$b \to \mu X$ (Combined)	273,281
c	$Z \to c\bar{c}$	137,740
	$t\bar{t} \to c$	404,753
	QCD $c\bar{c}$ (Merged)	90,441
	c (Combined)	1,076,054
$c \to \mu X$	$Z \to c\bar{c} \to \mu X$	47,986
	$t\bar{t} \to c \to \mu X$	17,715
	$c \to \mu X$ (Combined)	91,824
uds	$Z \to q\bar{q}$	152,590
	QCD Fake (Merged)	239,118
All	QCD All	2,086,603

8.8   Scale Factor Parametrization

Parametrizations of the tagging efficiency are used as tag rate functions (TRFs) and scale factors (SF) to simulate the performance of the tagging algorithm in simulations. The following functions are defined:

TRF$_b$: The efficiency to tag a b-jet in data.
TRF$_c$: The efficiency to tag a c-jet in data.
SF: The factor by which the b and c MC tagging efficiencies have to be multiplied by to obtain the data tagging efficiencies.

These functions are parametrized in terms of the E_T and η of the jets. As the use of the *SystemD* method requires high statistics to lead to stable solutions, it is not possible to extract a proper 2D parametrization. Instead, it is assumed

that the dependence on these variables can be factorized:

$$\varepsilon(E_\mathrm{T}, \eta) = \frac{1}{\varepsilon_\mathrm{all}} \cdot f(E_\mathrm{T}) \cdot g(|\eta|),$$

where ε_all is the efficiency for the entire sample, and

$$f(E_\mathrm{T}) = \frac{c}{1 + ae^{-bE_\mathrm{T}}}, \tag{16}$$
$$g(|\eta|) = d + e|\eta| + f|\eta|^2 + g|\eta|^3 + h|\eta|^4,$$

where $a - h$ are fit parameters.

The TRFs are directly measured for muonic b-jets in data and Monte Carlo, and for an inclusive b-jet sample in Monte Carlo only. A data/MC scale factor (SF) is determined as the ratio of muonic b-jet tagging efficiencies in data and Monte Carlo. The SF measures the effect on the tagging rate caused by the differences in tracking between data and MC. The profile of the SF in E_T and $|\eta|$ allows tagging rates to be corrected over the full E_T and $|\eta|$ phase space. The inclusive b-jet data TRF is calculated by multiplying the inclusive b-jet MC TRF by the SF. A similar procedure is used to determine the c-jet TRFs. **Shouldn't we say something about how well the factorization hypothesis works?**

8.9   Systematic Uncertainties

Uncertainties on the resulting efficiencies arise from the following sources: the *SystemD* calculations (due to uncertainties on the correction factors as well as limited data statistics); and the dependence of the TRFs on the MC samples, as well as possible imperfections in their chosen parametrization. These uncertainties are discussed in some detail below.

8.9.1   SystemD uncertainties

The correction factors α, β, κ_b and κ_cl (see Section 8.3) are evaluated on Monte Carlo. The effect of the uncertainty on each correction factor is evaluated by repeating the *SystemD* computations with the parametrization of that factor shifted by one standard deviation, while all other correction factors are fixed to their nominal values; the resulting changes in the computed efficiency are interpreted as systematic uncertainties. The effect of the choice of $p_\mathrm{T,rel}$ cut in the *SystemD* calculations is evaluated by varying the cut value between 0.3 GeV/c and 0.7 GeV/c. The total relative systematic uncertainty

associated with the *SystemD* input parameters is estimated by adding the individual contributions in quadrature, and varies between 1.3% and 1.7% for the different operating points.

For each bin in η and E_T, the *SystemD* systematic uncertainty for that bin is added in quadrature with the *SystemD* fit uncertainty. This yields an overall uncertainty with which the efficiency is known for each bin. This procedure is repeated for the *SystemD* efficiency of the entire sample. The combined uncertainties are then used to fit the parametrized curves in E_T and $|\eta|$, with the *SystemD* uncertainty folded into the statistical uncertainties.

The overall relative statistical errors are calculated by evaluating

$$\sigma_\mathrm{stat} = \frac{f^{+1\sigma}(E_\mathrm{T}) \cdot g^{+1\sigma}(|\eta|)}{\varepsilon_\mathrm{all}^{+1\sigma}} - \frac{f(E_\mathrm{T}) \cdot g(|\eta|)}{\varepsilon_\mathrm{all}}, \tag{17}$$

where σ_stat is the statistical error, f and g parametrize the efficiency as a function of jet E_T and $|\eta|$ and ε_all is the efficiency for the entire sample (see Eqn. 16); the $+1\sigma$ quantities are the corresponding fluctuations upward by one standard deviation. This is also repeated with the downward fluctuations; the larger deviation is taken as the statistical uncertainty.

The following comes from the old JLIP description:

As an illustration for the JLIP Medium working point, varying the input factors in *SystemD* uncertainties, the obtained results are summarized in Table 6. The overall relative uncertainty on the muon-in-jet b-tag efficiency is $\pm 1.4\%$ for all working points.

Table 6
Relative systematics on the b-tag efficiency with the *SystemD* method for the JLIP Medium working point. **This should be replaced with its NN equivalent, if deemed useful.**

Source	Relative uncertainty (%)
$\kappa_\mathrm{b} \pm 0.004$	± 0.9
$\kappa_\mathrm{cl} = 0.90 \pm 0.05$	$-$
$\beta = 1.016 \pm 0.003$	± 0.4
$\alpha = 1.0 \pm 0.8$	± 0.3
$p_\mathrm{T,rel} > 0.5$ to 0.9 GeV/c	± 0.9
Total	± 1.4

8.9.2 TRF parametrization and sample dependence

Both the parametrization and MC sample dependence systematic errors, which result from the use of TRFs derived from generic combined samples of MC b, c and muonic b jets, are quantified in one measurement. By comparing the percentage difference between the number of actual tags (N_{obs}) and the predicted number (N_{pred}) in various bins in E_{T} and η regions, effectively performing a closure test, a total error on the TRFs is determined from the spread of the percentage differences.

The closure tests are carried out for each of the MC samples used to construct the TRFs. The percentage differences are calculated in E_{T} bins in the CC, ICR and EC calorimeter regions, and the percentage differences are binned weighted by the number of actual tags in the region. The RMS of the resulting distributions are used to quantify the total error on each of the TRFs. The relative uncertainty determined by this method ranges from 1.2% for the lowest operating point to 3.5% for the tightest operating point for the inclusive b-jet TRF and from 2.4% to 4.0% for the inclusive c-jet TRF.

8.9.3 Total systematic uncertainty

The total systematic uncertainty is derived by adding the fit uncertainty of the parametrizations and the parametrization and sample dependence uncertainties in quadrature. The final systematic uncertainty ranges from 1.9% for the loosest operating point to 4.8% for the tightest operating point for TRF_{b}, from 2.8% to 5.2% for TRF_{c} and from 1.4% to 3.4% for the Data/MC scale factor.

The total errors, given by the statistical and systematic errors combined in quadrature, for the SF, TRF_{b} and TRF_{c} are shown in Fig. 19 for two operating points. The relative error increases rapidly at high η, due to limited statistics in that region and because the value of the scale factor drops rapidly for $|\eta| > 2$.

The following again comes from the old JLIP description:

As summarized in Table 7, systematics on the TRF_{b} and TRF_{c} tagging rate functions and on the SF_{b} scale factor include the *SystemD* uncertainties, the uncertainty on the jet (E_{T}, $|\eta|$) factorization hypothesis and the Monte-Carlo limited statistics. These contributions are quadratically added.

For JLIP, the TRF_{b} systematics range from $\pm3.3\%$ to $\pm2.0\%$ for the Ultra-Tight to SuperLoose working points. The TRF_{c} systematics are $\pm3.4\%$ in average for all working points. The SF_{b} systematics are $\pm1.7\%$ in average for all working points.
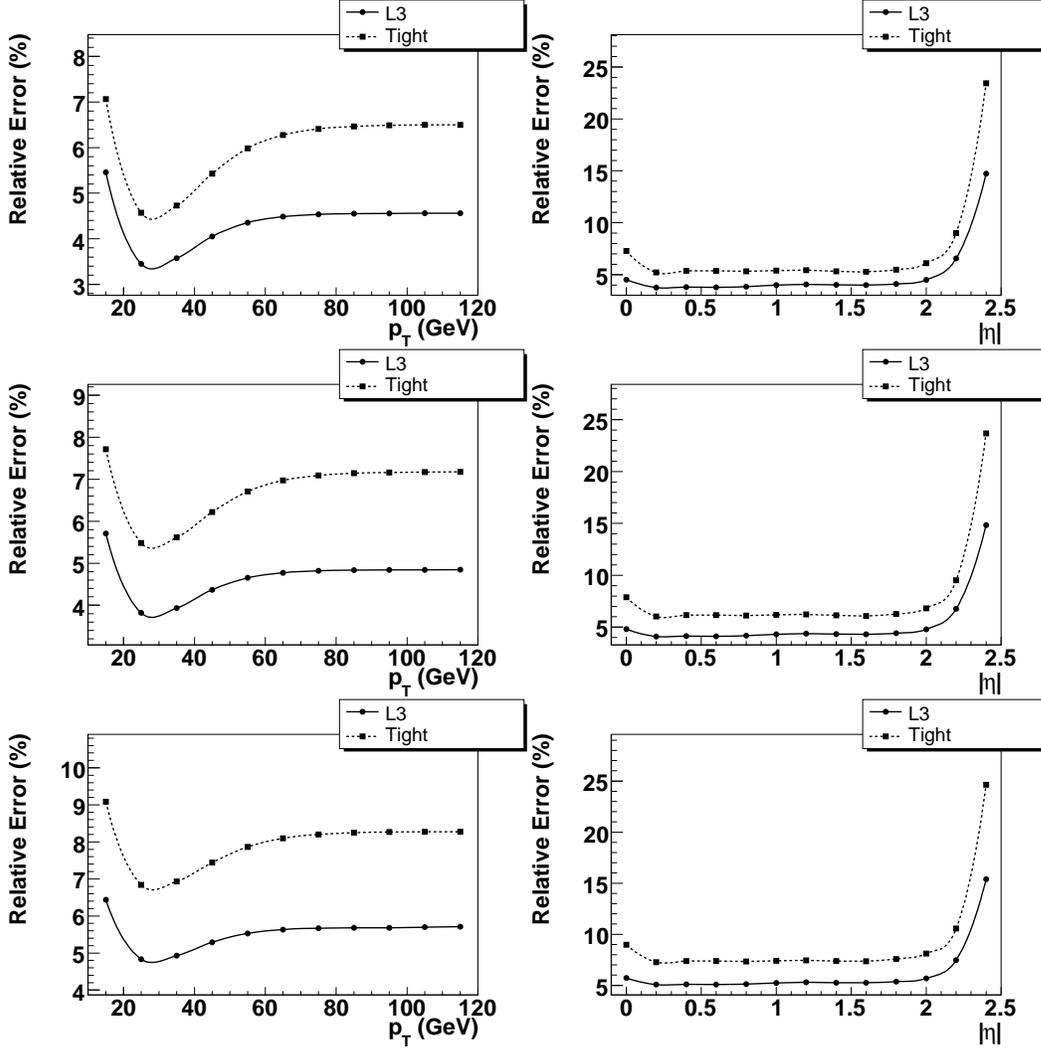
Fig. 19. The total relative error (combined systematic and statistical) for the Scale Factor (SF) (top), TRF_b (middle) and TRF_c (bottom) in terms of E_T (left) when $\eta = 1.2$ and η (right) when $E_T = 45$ GeV.

Table 7

Relative systematics on TRF_{b}, TRF_{c} and SF_{b}, for all JLIP working points.

JLIP cut	< 0.1%	< 0.3%	< 0.5%	< 1.0%	< 2.0%	< 4.0%
SystemD	±1.4%	±1.4%	±1.4%	±1.4%	±1.4%	±1.4%
factorization	±0.5%	±0.4%	±0.4%	±0.4%	±0.4%	±0.4%
inclusive b MC stat.	±2.8%	±2.4%	±2.1%	±1.6%	±1.3%	±1.1%
inclusive c MC stat.	±2.8%	±3.0%	±3.3%	±2.9%	±2.9%	±2.9%
b→ μX MC stat.	±1.0%	±0.8%	±0.7%	±0.7%	±0.7%	±0.7%
TRF_{b}	±3.3%	±2.9%	±2.6%	±2.3%	±2.1%	±2.0%
TRF_{c}	±3.3%	±3.4%	±3.7%	±3.3%	±3.3%	±3.3%
SF_{b}	±1.8%	±1.7%	±1.6%	±1.6%	±1.6%	±1.6%

9 Fake Rate Determination

The cornerstone of the determination of the light quark mistag rate (where "light" stands for uds-quark or gluon jets) or *fake rate* is the notion that neglecting long-lived particles such as the V^0s mentioned in Sect. 3.3, high-impact parameter tracks or displaced vertices reconstructed in light-flavour jets result from resolution effects. To good approximation, these effects should not depend on any jet direction.

This fact is exploited by using tracks with negative impact parameters (see Sect. 5 for the impact parameter sign convention used) and/or displaced vertices with negative decay lengths. Barring wrongly assigned negative impact parameter signs (which may occur whenever the jet and track are nearly aligned in azimuth, and which is important for long-lived particles), using such tracks and vertices should provide a reasonable estimate of the fake rate.

9.1 Data Sample

To minimize the impact of wrongly attributed impact parameter signs, the fake rate is determined in multijet data, which is dominated by generic QCD processes and hence has a low heavy flavour content. Two samples are used for this purpose:

- the EM sample: this sample consists of events selected by requiring at least one electron candidate with $p_T > 4$ GeV/c, and with low missing transverse energy, $\not{E}_T < 10$ GeV. At least one trigger unbiased in terms of b-tagging is required. Most of the electron candidates in reality are jets which deposit a large fraction of their energy in the EM section of the calorimeter. This may bias the sample compared to generic QCD processes, as the fraction of a jet's energy deposited through electromagnetic processes depends on the jet flavour. This bias is removed by only considering jets whose distance to the nearest identified EM cluster is larger than $\Delta R = 0.4$. After all these requirements, this sample contains 106 million taggable jets in 72 million events;
- the QCD sample: this sample consists of all events collected using pure jet triggers. It contains 154 million taggable jets in 249 million events. For the sake of consistency, jets in the vicinity of identified EM clusters are not considered either in this sample. Since trigger requirements should not bias such jets in this sample, the effect of this removal should be small, and will be evaluated below.

These two samples are combined for most purposes; their comparison allows to estimate the systematics associated with the choice of a particular sample.

Figure 20 compares the distributions of several kinematical variables between these samples and generic QCD events simulated using the PYTHIA [12] event generator: jet multiplicity per event, E_T, $|\eta|$ and track multiplicity per jet. As expected, the QCD sample is better described by the QCD simulation than the EM sample. **This last comment should be explained!?**
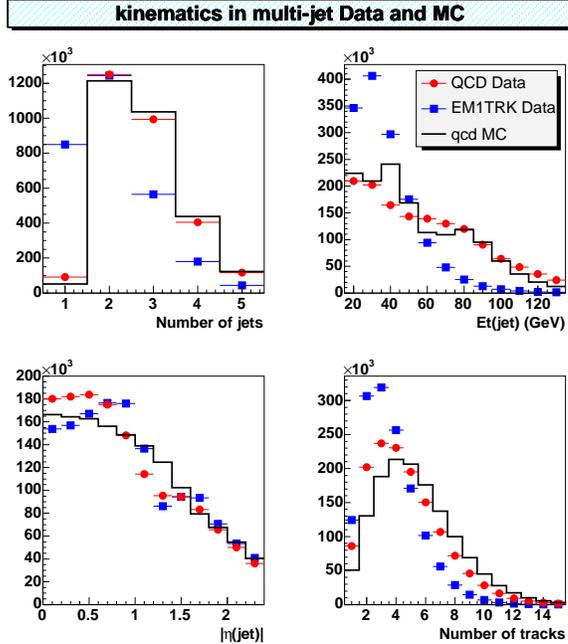
Fig. 20. Number of jets per event, jet E_T and $|\eta|$ and track multiplicity per jet in the QCD and EM data samples (dots) and QCD Monte-Carlo (solid line histogram, normalized to the number of entries). **NB this is a p14 plot and should be remade for p17!**

9.2 Negative Tag Rate

The use of negative impact parameter tracks and negative decay length displaced vertices is rather straightforward: the algorithms providing the NN input variables listed in Sect. 7.1 need only minor modifications in order to provide "negative" equivalents of these variables, called *Negative Tag* (NT) results in the following. The NN output is then simply recomputed using the above values rather than the original ones. In detail, the NN input NT results are computed as follows:

CSIP: The CSIP *Comb* variable is recalculated, using tracks with negative instead of positive impact parameter significance to obtain the "strong classifier" numbers of tracks N_{3s} and N_{2s} (see Sect. 6).

JLIP: The Jet Lifetime Probability $\mathcal{P}_{\mathrm{JLIP}}$ is recomputed using only tracks with negative rather than positive impact parameter significance.

SVT: In this case, no additional computation is necessary. Instead of the highest (positive) decay length significance the most negative decay length significance displaced vertices (for both algorithm versions discussed in Sect. 4) are used to supply the SVT-related NN variables.

Like the b-jet efficiency, the fake rate and negative tag rate are parametrized as a function of a jet's kinematical $(E_\mathrm{T}, |\eta|)$ variables. However, in contrast to the efficiency, it is found **(is that on the basis of MC??)** that the dependence of the negative tag rate on jet E_T and $|\eta|$ cannot well be considered to factorize into a dependence on E_T multiplied by a dependence on $|\eta|$. Instead, it is parametrized as a function of jet E_T in three regions: $0 < |\eta| < 1.2$ (CC), $1.2 < |\eta| < 1.8$ (ICR) and $1.8 < |\eta| < 2.4$ (EC). In each region, the E_T dependence is parametrized using a quadratic polynomial.

Here should follow some plots: comparison of NT rates and corresponding parametrizations in the three regions (could be two plots, one for a loose OP and one for a tight OP??), as well as accompanying text (remark on qualitative behaviour?).

9.3   Corrections

The Negative Tag rate is not a perfect approximation of the fake rate. Corrections for the following effects are applied:

- the presence of heavy flavour jets increases the NT rate, primarily due to tracks that originate from the decay of long-lived particles and that are (mistakenly) assigned a negative impact parameter sign. As no method is available to estimate this effect on real data, simulated (generic QCD) events are used instead. This results in a correction factor $F_{hf} = \varepsilon^-_{\mathrm{QCD,light}}/\varepsilon^-_{\mathrm{QCD,all}}$, *i.e.*, the ratio of Negative Tag rates with and without the presence of heavy flavour jets in these simulated events;
- the V^0 removal algorithm (see Sect. 3.3) which is not fully efficient, so that some contribution from long-lived particles like K_S, Λ etc. remains. Most of the resulting tracks will correctly be assigned positive impact parameters, and the NT rate is affected less by their presence than the fake rate. Also this effect is estimated using simulated events, leading to a correction factor $F_{ll} = \varepsilon^+_{\mathrm{QCD,light}}/\varepsilon^-_{\mathrm{QCD,light}}$, *i.e.*, the ratio of the fake rate and the light-flavour NT rate in simulated events.

Finally, the fake rate is estimated as the NT rate corrected for the above effect:

$$\varepsilon_{\mathrm{light}} = \varepsilon^-_{\mathrm{data}} \cdot F_{hf} \cdot F_{ll}. \tag{18}$$

As an illustration, for the JLIP Medium working point (see Sect. 5), the jet E_T dependence of F_{hf}, F_{ll}, of the negative tag rate $\varepsilon_{\mathrm{data}}^-$ and of the estimated light quark tagging efficiency $\varepsilon_{\mathrm{light}}$ are presented in Fig. 21.
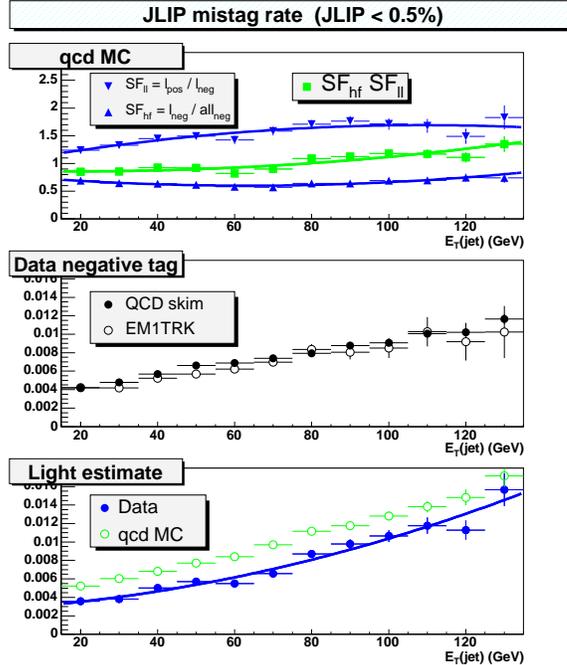
Fig. 21. Jet E_T dependence of (top) F_{hf} and F_{ll} in QCD simulation, (middle) the negative tag rate in multi-jet data and (bottom) the estimated light quark tagging efficiency for the JLIP Medium working point. **This figure is p14 and should be replaced with a p17 equivalent...**

When the jet energy gets higher, the multiplicity of long lived particles and their average decay length increase, giving larger impact parameters to their decay products. The charged particles get closer and closer from the jet axis, leading also to a larger number of wrong sign impact parameter assignments. These effects contribute to explain why the negative and positive tag rates increase with E_T for light quark jets. The data behaviour is similar to that predicted by the Monte-Carlo simulation, however the ratio of the data to Monte-Carlo light quark tagging efficiencies is found to be of about 0.7.

9.4 Systematic Uncertainties

The use of a particular sample (in this case, the combination of QCD and EM multijet samples) to provide a "universal" estimate of the NT rate needs to be validated. To this end, the ratio of the NT rates as measured in the separate QCD and EM samples is determined as a function of the kinematical variables, **and shown in Fig. (some figure) for the (some OP) operating point(s).**

And here should follow the figure(s).

A corresponding systematic uncertainty is calculated from a constant fit to the EM/QCD NT rate ratio. Half the difference between the fit value and unity is taken as the systematic error, or if the ratio is consistent with unity within the fit uncertainty scaled by $\sqrt{\chi^2/N_{\mathrm{DF}}}$, this scaled fit uncertainty is taken as the error. The relative uncertainty ranges from 0.1% to 0.7% for the different operating points and detector regions.

In addition to the difference between the QCD and EM samples, the effect of removing jets in the vicinity of EM clusters in the QCD sample needs to be taken into account. The NT rate in the QCD sample with the jets removed is slightly lower than in the full QCD sample. The effect is small, ranging from 0.2% for the loosest and almost 1% for the tightest operating point, and does not depend on jet E_{T}. A systematic error is assigned in the same way as that for the difference between the EM and QCD skims, and ranges from 0.2% to 1.0% for the different operating points and detector regions. **Do we need a plot here, too?**

To test the parametrization of the NT rate in the three $|\eta|$ regions a comparison is made between the number of tags found by the tagger and its prediction from the parametrized NT rate. A systematic uncertainty is again calculated from a constant fit to the ratio of the actual and predicted number of tags, following the same procedure as the EM/QCD sample comparison. The systematic uncertainty ranges from 0.06% to 0.7% for the different operating points and detector regions.

The F_{hf} correction factor depends on the assumed b- and c-fractions in the multi-jet data sample. In turn, these depend on the cross sections for QCD heavy flavour production, which, until a few years ago, exhibited a factor of two difference between experimental measurements performed at the Tevatron collider [13] and corresponding NLO theoretical predictions (see *e.g.* [14]). Recent developments [15] indicate that the discrepancy is presumably largely due to an unsatisfactory treatment of fragmentation effects in the experimental measurements, but a precise comparison is still lacking. To estimate the uncertainty on F_{hf}, therefore, the fractions of b(c) jets are varied from their default values of 4.6% (2.6%) by 20% (relative). Given that the individual beauty and charm production mechanisms are very similar, these fractions are varied coherently. **We should say here something about the effect of this source of uncertainty!**

The total uncertainty on the fake tag rate is given by adding in quadrature the systematics contributions (as discussed above) for the appropriate region to the statistical uncertainty, estimated as the difference between the fake tag rate central value and the one standard deviation fit curves. The dominant

48

contribution is the systematic one, although the statistical error on the fits has an increasing contribution as the operating point becomes tighter. The combined relative uncertainty ranges from 3.7% (for the loosest operating point) to 9.2% (for the tightest operating point) in the CC region, from 3.0% to 9.5% in the ICR region, and from 2.0% to 5.6% in the EC region. **To be added here: text introducing Table 8.**

Table 8
Relative systematics on the mistag rates, for all JLIP working points. **NB: obsolete, to be replaced by a p17 equivalent**

JLIP cut	$< 0.1\%$	$< 0.3\%$	$< 0.5\%$	$< 1.0\%$	$< 2.0\%$	$< 4.0\%$
$QCD/EM1TRK$	$\pm 5.9\%$	$\pm 5.1\%$	$\pm 3.6\%$	$\pm 2.6\%$	$\pm 2.2\%$	$\pm 1.7\%$
factorization	$\pm 2.2\%$	$\pm 1.5\%$	$\pm 2.1\%$	$\pm 1.0\%$	$\pm 1.1\%$	$\pm 0.8\%$
c-rate	$\pm 2.3\%$	$\pm 2.2\%$	$\pm 2.1\%$	$\pm 1.9\%$	$\pm 1.5\%$	$\pm 1.2\%$
b-rate	$\pm 7.5\%$	$\pm 5.5\%$	$\pm 4.6\%$	$\pm 3.5\%$	$\pm 2.5\%$	$\pm 1.7\%$
Total syst.	$\pm 11.6\%$	$\pm 9.4\%$	$\pm 7.9\%$	$\pm 6.1\%$	$\pm 4.7\%$	$\pm 3.5\%$

10   Performance of the NN tagger (b-Efficiency Measurement)

The Run II data performance of the NN and JLIP taggers as measured on data including full statistical and systematic errors on the measurement is shown in Fig. 22 for all jets and also for jets with $\eta < 0.8$ and $E_\mathrm{T} > 30$ GeV. The NN tagger demonstrates considerable improvement over the JLIP tagger, which is the most-used tagger before the NN tagger is developed, for all operating points.
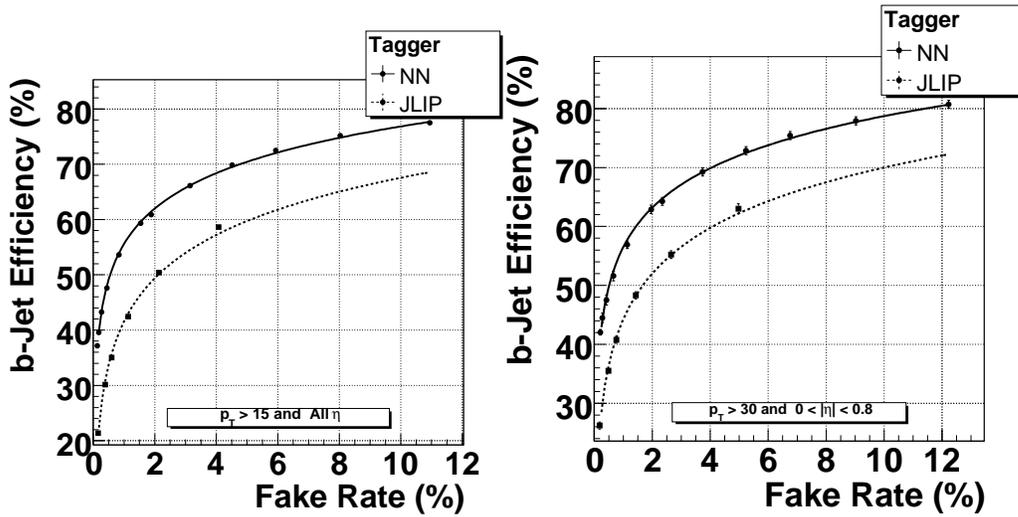
Fig. 22. Performance profile of NN and JLIP taggers on the μ+jets and multi-jet samples. The top plot is for all jets and the bottom plot for jets with $\eta < 0.8$ and $E_\mathrm{T} > 30$ GeV. The error on the plots represents the total uncertainty, statistical and systematic, on the performance measurements. The NN tagger demonstrates large performance gains over the JLIP tagger, with increases in efficiency of up to 50% for a fixed fake rate. Fake rates are typically reduced to between a quarter and a third of their value for a fixed signal efficiency.

References

[1] W.-M. Yao, et al., Review of Particle Physics, Journal of Physics G 33 (2006) 1.

[2] F. Abe, et al., Observation of Top Quark Production in $\bar{p}p$ Collisions, Phys. Rev. Lett. 74 (1995) 2626–2631.

S. Abachi, et al., Observation of the Top Quark, Phys. Rev. Lett. 74 (1995) 2632–2637.

[3] V. Abazov, et al., The Upgraded DØ Detector, Nucl. Instrum. Meth. A565 (2006) 463, FERMILAB-PUB-05/341-E.

[4] V. Abazov, et al., Simultaneous measurement of the ratio $B(\text{t} \to W\text{b})/B(\text{t} \to Wq)$ and the top quark pair production cross section with the D0 detector at $\sqrt{s} = 1.96$ TeV, arXiv:0801.1326 [hep-ex], accepted by Phys. Rev. Lett. (2008).

[5] R. Frühwirth, Application of Kalman filtering to track and vertex fitting, Nucl. Instrum. Meth. A262 (1987) 444–450.

[6] P. Hough, Method and Means for Recognizing Complex Patterns, U.S. Patent no. 3,069,654 (1962).

R. Duda, P. Hart, Use of the Hough Transformation to Detect Lines and Curves in Pictures, Comm. ACM 25 (1972) 449–456.

[7] R. Frühwirth, P. Kubinec, W. Mitaroff, M. Regler, Vertex reconstruction and track bundling at the LEP collider using robust algorithms, Comput. Phys. Comm. 96 (1996) 189–208.

J. D'Hondt, P. Vanlaer, R. Frühwirth, W. Waltenberger, Sensitivity of robust vertex fitting algorithms, IEEE Trans. Nucl. Sci. 51 (2004) 2037–2044.

[8] Fermilab, Run II jet physics, FERMILAB-CONF-00-092-E.

[9] Research directions for the decade. Proceedings, 1990 Summer Study on High-Energy Physics, Snowmass, USA, June 25 - July 13, 1990, singapore, Singapore: World Scientific (1992) 808 p.

[10] D. Buskulic, et al., Phys. Lett. B313 (1993) 535.

G. Borisov, C. Mariotti, Nucl. Instrum. Meth. A372 (1996) 181.

[11] http://www.root.cern.ch/. The training method chosen is the Broyden-Fletcher-Goldfarb-Shanno (BGFS) method.

[12] T. Sjöstrand, et al., High-energy-physics event generation with PYTHIA 6.1, Comput. Phys. Comm. 135 (2001) 238–259.

[13] B. Abbott, et al., Cross section for b jet production in $\bar{p}p$ collisions at $\sqrt{s} =1.8$ TeV, Phys. Rev. Lett. 85 (2000) 5068.

D. Acosta, et al., Measurement of the B^+ total cross section and B^+ differential cross section $d\sigma/dp_T$ in $\bar{p}p$ collisions at $\sqrt{s} =1.8$ TeV, Phys. Rev. D65 (2002) 052005.

[14] S. Frixione, M. L. Mangano, P. Nason, G. Ridolfi, Heavy-quark production, Adv. Ser. Direct. High Energy Phys. 15 (1998) 609–706.

[15] M. Cacciari, P. Nason, Is there a significant excess in bottom hadroproduction at the Tevatron?, Phys. Rev. Lett. 89 (2002) 122003.