Turk J Phys
26 (2002) , 81 – 88.

A Method for Identifying Gluon Jets with High Purity in 3-jet Hadronic Z Decays

Ayda BEDDALL
Engineering Physics Department, University of Gaziantep,
Gaziantep 27310-TURKEY
e-mail: beddall@gantep.edu.tr

Received 24.04.2001

Abstract

A method for identifying gluon jets with very high purity is presented and the performance evaluated. Quark and gluon jets from Monte Carlo simulated Hadronic Z decays are identified by energy ordered jets and applying heavy flavour tagging. It is shown that quark and gluon jets can be identified with a purity of about 98%.

Key Words: Three-Jet; Gluon purity; B-tag; Z Boson.

1. Introduction

In the reaction $e^+e^- \rightarrow q\bar{q}g$, three hadronic jets are produced, the third jet resulting from hard gluon radiation. The identification of gluon jets allows for the study of differences between quark and gluon jets; an example of such an analysis and the methods employed to acheive this can be found in [1]. Methods employed to identify gluon jets on an event-by-event basis involve the utililization of jet-finding algorithms and high resolution vertex detectors. In this study a method for identifying gluon jets to a very high purity is presented. Quark and gluon jets are identified by energy ordering the jets and heavy flavour tagging. For the determination of the performance of method, hadronic decays of the Z boson are simulated using the JETSET Monte Carlo [2]. Detector resolution plays an important role and so the Monte Carlo events are passed through a full detector simulation and reconstruction program for the ALEPH detector; the ALEPH detector is described in detail in [3] and [4].

2. 3-Jet Event Selection

Three-jet events are selected from hadronic decays of the Z by applying a jet finding algorithm and requiring that the number of jets is three. Additionally, a symmetric Y-shape event topology is selected to avoid jet energy biases for the comparison of differences between quark- and gluon-induced jets.

2.1. Jet-Finding Algorithms

Jet finding algorithms in e^+e^- annihilation are based on the concept of hierarchical clustering, combining 'nearby' particles pairwise to form pseudoparticles. This continues iteratively until the event consists of a few well-separated pseudoparticles, which are the output jets. Mass and momentum cut-offs are used to define what is meant by 'nearby'.

Basically, a jet clustering algorithm of the types most commonly used in e^+e^- annihilation is defined according to the following procedure:

1. Define a jet resolution parameter, y_{cut}.
2. Calculate the corresponding separation y_{kl} for every pair of the particles h_k, h_l.
3. If y_{ij} is the smallest value of y_{kl} calculated in (2) and ($y_{ij} < y_{cut}$), combine (p_i, p_j) into a single 'pseudoparticle' momentum p_{ij} according to a recombination scheme which is explained below.
4. Repeat this procedure from step (2) until all pairs of particle and/or pseudoparticles have $y_{kl} > y_{cut}$. The remaining objects at this stage are called jets. The number of pseudoparticles gives the number of the jets in the event.

The scaled invariant mass squared algorithm, JADE [5], can be used to construct the jets in hadronic events by merging particles together adding their 4-vectors to form pseudo-particles. In this algorithm the invariant mass of every pair of hadrons in the final state is calculated. If any are less than a certain fraction y_{cut} of the total centre of mass energy, s, then the momenta of the pair with the lowest invariant mass are added together. This combined momenta is considered to be that of a single 'particle'. The invariant masses are recalculated and the combining procedure continued until none of the invariant masses are less than y_{cut}. For all pairs of particles k and l the scaled invariant mass squared y_{kl} is calculated by

$$y_{kl} = M_{kl}^2/E_{vis}^2, \tag{1}$$

where E_{vis} is the visible energy of the event computed as the sum of the input particle energies. The squared invariant mass M_{kl}^2 for the JADE algorithm is

$$M_{kl}^2 = 2E_kE_l(1 - cos\theta_{kl}), \tag{2}$$

where E_k and E_l are the energies of k and l tracks which are separated by an angle θ_{kl}.

If their scaled invariant mass is smaller than y_{cut} ($y_{kl} < y_{cut}$), the tracks are merged. The loop is run over the new list of tracks which has lost 2 particles and gained the merged pair. This procedure is repeated until all the remaining y_{kl} exceed a threshold y_{cut} value. When none of the remaining pairs have a low enough mass the track list contains a set of merged tracks called jets.

Alternatively, in the DURHAM algorithm the scaled transverse momentum is defined as

$$y_{kl} = \frac{(M_T^2)_{kl}}{s} \tag{3}$$

$$(M_T^2)_{kl} = \min(E_k^2, E_l^2)\ 2(1 - \cos\theta_{kl}), \tag{4}$$

where E_k and E_l are the energies of the particles or jets and θ_{kl} is the angle between their momenta.

At small θ_{kl}, y_{kl} measures the minimal transverse momentum squared of one jet relative to the other, while at large angles the definition is such that y_{kl} always increases with angle. y_{kl} is first calculated for every pair of final state particles (k, l). Then the two particles (i, j) with the smallest value of y_{kl} are combined and replaced by a 'pseudoparticle' with 4-momentum (P_{ij}) if their y_{ij} is smaller than the specific resolution parameter y_{cut}. This procedure is repeated until all pairs of objects (particle and/or pseudoparticles) have $y_{kl} > y_{cut}$. The jet multiplicity n is defined in such a way that jets i and j with energies E_i, E_j at relative angle θ_{ij} are resolved if $y_{ij} > y_{cut}$.

2.2. Y-shape Event Selection

A symmetric Y-shape event topology is selected to avoid biases when comparing differences between quark- and gluon-induced jets, arrising from the varying energies of quark and gluon jets. Y-shape events are defined as those where the angles between the most energetic and each of the remaining jets are 150±10 degrees. The two lower energy jets overlap in energy, this allows the comparison of quark and gluon jets at the same energy to be made. Another advantage of selecting Y-shape events is that for these events the most energetic jet can be assumed to be a quark jet with a very high purity. This gives a higher probability of labelling the remaining jets correctly.

The event selection described in the above sections yields an event sample which in this study is named S_{mix}, indicating that it contains a mixed quark flavour contribution. Later, the selection of a sub-set of these events will be described yielding an event sample named S_{btag} where a double b-tag is applied to select heavy flavour events where two jets are identified as originating from b quarks.

3. Jet Identification

To compare the properties of quark and gluon jets it is necessary to be able to distinguish the two types. Two methods for acheiving this: energy ordering of jets, and heavy flavour tagging, are described in the following sections.

3.1. Energy Ordering

Because of the nature of the bremsstrahlung-like spectrum of gluons in $e^+e^- \rightarrow q\bar{q}g$ decays the least energetic jet (called jet-3) should have the highest probability of being a gluon jet. The energy ordering of jets can therefore be used to separate samples of quark and gluon jets. The jets within each event are energy ordered with the most energetic jet labelled jet-1, the second most energetic labelled jet-2 and the least energetic jet labelled jet-3. The energy of each jet can be calculated in two different ways, giving the visible energy E_{vis} or the calculated (kinematic) energy E_{cal}.

The visible energy of a jet can be described as the sum of the energy of all energy flow particles. The visible energy E_{vis} of a jet is defined simply by

$$E_{vis} \;=\; \sum_{i=1}^{n_j} E_{ij}, \tag{5}$$

where n_j is the number of energy flow particles in jet j (j=1,2,3), and E_{ij} is the energy of the i^{th} particle in jet j. The jets are ordered according to their calculated energies such that $E^1_{vis} > E^2_{vis} > E^3_{vis}$. As most particles in Hadronic Z decays are pions then energy flow is calculated assuming the pion mass for charged particles and zero mass for all neutrals (photons from decays of neutral pion).

The calculated energy of every jet in an event is determined by using the formula:

$$E^i_{cal} = \frac{\sin\theta_{jk}}{\sin\theta_{12} + \sin\theta_{23} + \sin\theta_{31}}\sqrt{s} \qquad (i,j,k \;\; cyclic), \tag{6}$$

where i, j, k represent the jet numbers with their inter-jet angles θ_{ij}, θ_{ik}, θ_{jk}, and $\sqrt{s}$ is the center of mass energy. The momentum sum of the three jets should be equal to zero and also the sum of the visible energy of three jets in every event should be equal to the center-of mass energy. The jets are then ordered according to their calculated energies such that $E^1_{cal} > E^2_{cal} > E^3_{cal}$. Calculating energy in this way improves jet resolution by reducing the effect of reconstruction inefficiencies and track resolution.

3.2. Heavy Flavour Tagging

Identifying jets initiated by heavy quarks improves considerably the accuracy of the identification of gluon-jets. A sub-set of the events selected so far can be selected by tagging heavy flavour events, a double b-tag [6] is applied so that events are only selected if two jets are identified as originating from a b- or $\bar{b}$-quark, the remaining jet being identified, to a high accuracy, as a gluon-induced jet. The long life and large mass of the heavy quarks cause their decay products to have large impact parameters, allowing a separation from other quarks. Hadrons containing b quarks can be separated from the other hadrons by using tracks with large impact parameter or tagging the secondary vertex. The advantages of the impact parameter technique over explicitly reconstructing a secondary vertex are: it is conceptually simple, tertiary vertices occur as b hadrons typically decay into c hadrons, the tracks from these vertices contribute to the statistical power of the impact parameter method. The track impact parameter method for heavy flavour tagging of events is described next.

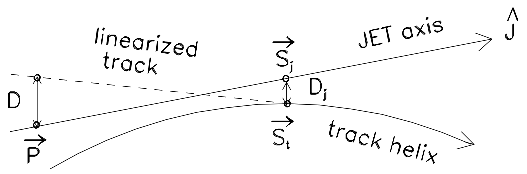

Figure 1. Impact parameter magnitude and sign definition in 3 dimensions.

The signed impact parameter of every track in an event produced in a hadronic Z decay is computed. The impact parameter is constructed from 3-dimensional information. The original b hadron direction is obtained from the direction of the jets. The primary vertex position is calculated from the jet direction, the beamspot position, and all tracks in the event. The impact parameter magnitude and sign definition in three-dimensions is shown in Figure 1. Point $\vec{P}$ shows the position of the primary vertex. $\hat{J}$ is the direction of the b hadron, as approximated by the jet direction. The circular arc represents a track, assumed here to have originated from the b decay point. $\vec{S}_t$ is the point on the track where it comes closest to the jet axis. $\vec{S}_j$ is an approximation of the b decay point. The track is linearized at $\vec{S}_t$, and the impact parameter magnitude D is defined as the distance of closest approach between the linearized track and the primary vertex. The physics principle of signing is that the decay point of the b hadron must lie along the flight path of the hadron. The sign is then defined as positive when the vector $\vec{S}_j - \vec{P}$ lies in the same direction as the jet direction $\hat{J}$. Detector resolution generates a random sign for tracks which originate from the primary vertex. Since only positive tracks are used in the tag, this reduces by 50% the background from tracks without lifetime. The negative signed tracks form a control sample which can be used to measure the resolution. The signed impact parameter $\tilde{D}$ is defined for each track by

$$\tilde{D} = \frac{(\vec{S}_j - \vec{P}) \cdot \hat{J}}{|(\vec{S}_j - \vec{P}) \cdot \hat{J}|} \, D \tag{7}$$

where $\tilde{D}/\sigma_D$ defines the estimated statistical significance of the measured impact parameter.

Tracks with positive impact parameter are used to obtain the lifetime information. The resolution function σ_D can be measured from the negative impact parameters. A probability function is defined for the tracks to represent the probability that the measured positive $\tilde{D}/\sigma_D$ is consistent with the fit to the negative significance of $\tilde{D}/\sigma_D$ (the probability that the particle originated at the primary vertex). The track probability is then defined as the result of this function, and used to tag the b hadrons.

Figure 2 shows the track probability distribution for b, c, and u d s quark events. The figure shows for b events that there is a large spike near zero on the positive side which is the lifetime signal in the events. A similar, though much smaller, spike is seen in charm events coming from the lifetime of the c hadrons. The negative side of the distributions represents the tracks with negative $\tilde{D}$. For light flavour events the primary quarks contribute no lifetime, so that the spectrum between 0 and 1 is expected to be flat as seen on the negative side of the spectrum. The tracks from the b decay must be close to jet axis (i.e associated with a b hadron). Therefore a further cut on D_j, which is the distance of closest approach of the track to the axis (see Figure 1), is applied to reduce the number of included badly-measured tracks. Some background comes from poorly reconstructed tracks, such tracks have a large value of D and fake the b signal. The background due to these track is reduced by an appopriate cut on D_j. The physics principle applied is that, even if a track comes from a b hadron decay and thus has a large impact parameter at the origin, it must still come close to some point on the hadron trajectory. Conversely poorly reconstructed tracks have a large impact parameter subject to no such constraint. Tracks from b decays should not be too far from the primary vertex since the mean decay length of a b meson is in the order of millimeters.

The number of tracks with large impact parameters can be used to distinguish b hadrons from the others. Beauty hadrons typically produce about 5 charged tracks that contain lifetime information in their impact parameters. An event probability P_{Event} is then obtained by using the individual track probabilities. The event probability is defined as the probability that any group of N tracks without lifetime produce

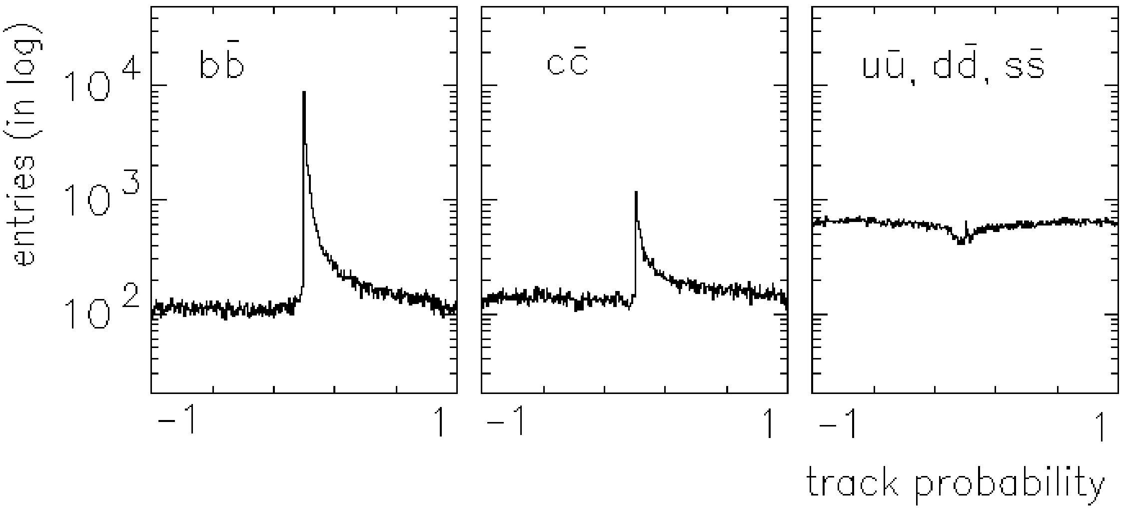

Figure 2. Track probability distribution. The plots show Monte Carlo events divided into $b\bar{b}$, $c\bar{c}$, and light flavours respectively.

Table 1. Event tag efficiency and purity for $b\bar{b}$ events for different event tag values using Monte Carlo simulated events.

Prob. cuts	0.0001	0.001	0.003	0.005	0.01	0.03	0.05	0.1
b purity (%)	97	93	90	88	84	76	71	62
ε_b (%)	40	53	60	64	69	77	81	86

the observed values of track probabilities equally or more unlikely. More information on the mathematical calculation of the event probability can be found in reference [7]. Event probability distributions for $b\bar{b}$, $c\bar{c}$, and light (uds) quarks using Monte Carlo simulated events are shown in Figure 3. As expected, in $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ events the event probability distribution is flat indicating no lifetime in these events. The event cut which is used for b event selection is shown on the linear scale plot (inset).

To determine a suitable value for the event probability cut, purity and efficiency distributions are investigated, where purity is defined as

$$\frac{number\ of Z \rightarrow\ b\bar{b}\ events\ selected}{the\ total\ number\ of\ events\ selected} \tag{8}$$

and efficiency is defined as

$$\frac{number\ of Z \rightarrow\ b\bar{b}\ events\ selected}{the\ total\ number\ of\ Z \rightarrow\ b\bar{b}\ events} \tag{9}$$

Figure 4 shows the purity and efficiency distributions as a function of the *log* of the cut variable for event tags in $Z \rightarrow q\bar{q}g$ events; efficiencies for selecting charm and light flavour events is also shown.

Table 1 shows the numerical values for $b\bar{b}$ purity and efficiency for different values of the event probability cut. In this study, events are selected if the event probability is less than 0.001 which gives high b event purity, 93%, with a reasonable efficiency, 53%. After this cut has been applied ~11% of the 3-jet events are accepted.

For the identification of gluon jets within each selected heavy flavour event, the b-tag procedure is repeated for each individual jet yielding jet probabilities P_{jet1}, P_{jet2} and P_{jet3} for the three energy-ordered

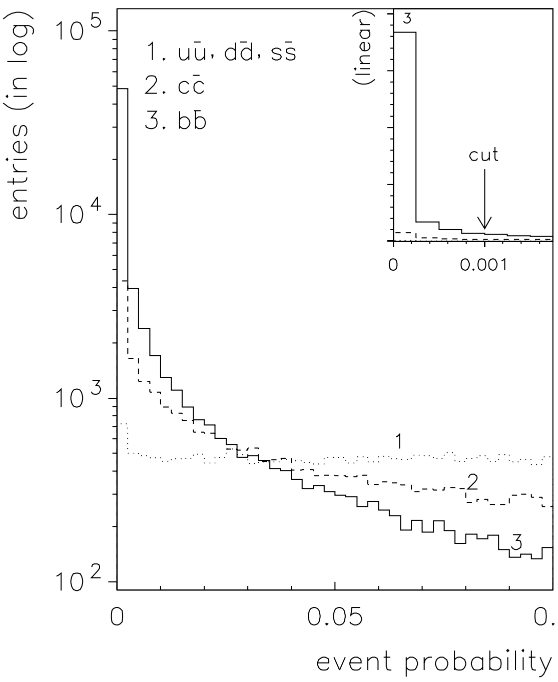

Figure 3. Event probability distribution for $b\bar{b}$, $c\bar{c}$, and light (uds) quarks using Monte Carlo simulated events. The cut which is used for the b event selection is shown on the linear scale plot (inset). This cut value gives high purity with acceptable efficiency of b events.

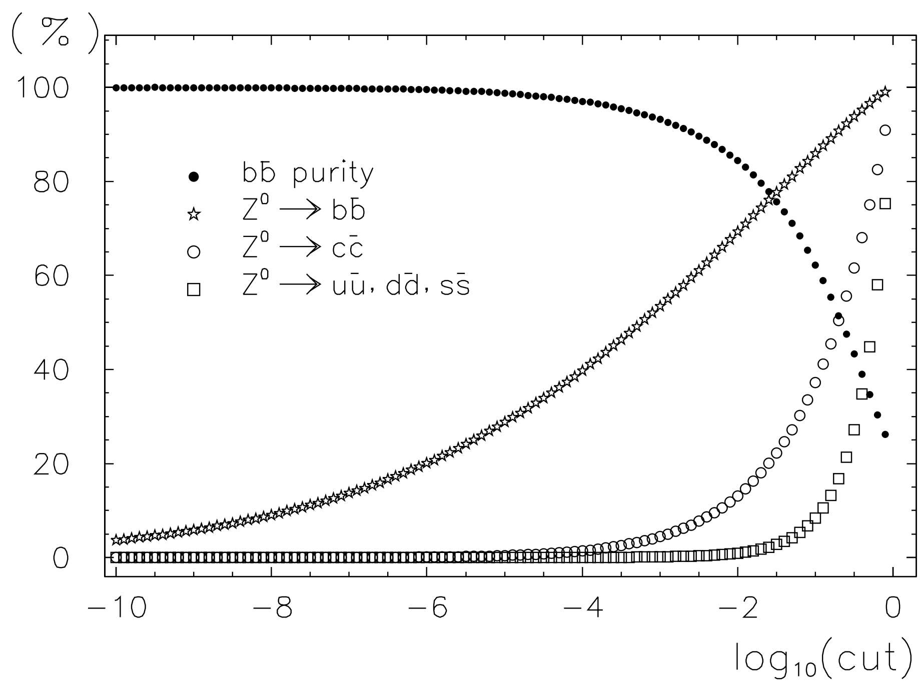

Figure 4. Purity and efficiency for event tagging with different quark species, as a function of the *log* of event probability cut.

Table 2. Number of correctly tagged, wrongly tagged and ambiguous events for jet1 (quark or antiquark), jet2 (antiquark or quark), and jet3 (gluon).

samples	total matched events	jet	correctly matched	wrongly matched	ambiguous
S_{mix}	7662	1	7442	220	98
		2	5593	2069	
		3	5210	2452	
S_{btag}	370	1	365	5	2
		2	363	7	
		3	362	8	

jets, respectively. Events are selected where jet-3 is determined, to a very high probabilty, to be a gluon jet by applying the following jet probability cuts:

- $P_{jet1} < 0.01$, to ensure that jet-1 contains a b- or $\bar{b}$-quark
- $P_{jet2} < 0.1$, to ensure that jet-2 contains a $\bar{b}$- or b-quark
- $P_{jet3} > 0.2$, to ensure that jet-3 does not contain a heavy flavour quark, and is therefore a gluon-induced jet.

After these cuts have been applied approximately one-third of the heavy flavour events remain yeilding a double b-tagged sample S_{btag}.

4. Results

The performance of the above methods of identifying gluon jets is measured by comparing the relative event selection efficiency and quark/gluon purities for the S_{mix} and S_{btag} event samples. For the purity calculation events are classified as correctly tagged, wrongly tagged or ambiguous events. Events are named as 'correctly tagged' events if the lowest energy quark jet on the parton level matches the second jet on the reconstructed level, and if the gluon jet on the parton level matches the third jet on the reconstructed level. If the lowest energy quark jet on the parton level matches the third jet on the reconstructed level, or the gluon jet on parton level matches the second jet on the reconstructed level then the event is named as 'wrongly tagged'. Ambiguous events are defined as:

- events with no gluon decays, such that there are only $q\bar{q}$ at the parton level.
- events where any two jets are clustered into the same jet.
- events where a minimum matching angle between the parton level and detector level is greater than 40 degrees.

The results for the two event samples are shown in Table 2. The purity of quark and gluon jets are obtained by using the formula:

$$P = \frac{N_{cor} + N_{amb}/2}{N_{cor} + N_{wro} + N_{amb}} \tag{10}$$

where N_{cor} and N_{wro} are the number of selected $Z \to q\bar{q}g$ events correctly and wrongly matched to the parton level events, respectively, and N_{amb} is the number of the ambiguous events. The purity calculation includes the ambiguous events, assuming that on average half of these are tagged correctly. The associated systematic error is calculated using

Table 3. Purities of jet1 (as quark or antiquark), jet2 (as antiquark or quark), and jet3 (as gluon) for S_{mix} and S_{btag} samples. The first error is statistical and the second systematic.

samples	purity %		
	jet1	jet2	jet3
S_{mix}	96.53±0.21±0.63	72.71±0.51±0.64	67.78±0.53±0.68
S_{btag}	98.39±0.65±0.27	97.85±0.75±0.27	97.58±0.80±0.27

$$P_{max} = \frac{N_{cor} + N_{amb}}{N_{cor} + N_{wro} + N_{amb}} \tag{11}$$

$$P_{min} = \frac{N_{cor}}{N_{cor} + N_{wro} + N_{amb}} \tag{12}$$

where, in the first equation, the maximum purity is calculated by supposing that the ambiguous events are correctly tagged, while in the second equation all of the 'ambiguous' events are considered as tagged wrongly. The statistical errors on the purity of jets are calculated using

$$\Delta P_{stat} = \sqrt{\frac{P\ (1-P)}{N_{evnt}}}, \tag{13}$$

where N_{evnt} is the number of the selected events.

The results for the jet identification purity for the two event samples is shown in Table 3.

5. Summary and Conclusion

In this study a method for the identification of quark and gluon jets to a very high purity is described. Three-jet events are selected from Monte Carlo Hadronic Z decays by employing a jet finding algorithm. Energy ording the jets provides the first level of jet identification where the least energetic jet can be assumed to be a gluon jet with a purity of about 68%. While the highest energy jet can be assumed to be a quark induced jet with a purity of about 96%, the quark purity of the second jet is only 73%.

The accuracy of jet identification improves greatly by b-tagging heavy flavour events and applying a double b-tag to jets, selecting a sub-sample of events where both the first and second jets are identified as heavy flavour jets. While this greatly reduces the number of selected 3-jet events to only about 4%, quark and gluon jets are identified with a very high purity of about (98 ±1)% thereby allowing comparisons of quark and gluon jets with little uncertainty due to impurities. Statistical uncertainties due to the very low efficiency for selecting b-tagged 3-jet events can be small given that LEP experiments have recorded millions of hadronic Z decays.

References

[1] D. Buskulic et al. (ALEPH Collab.), *Phys. Lett. B* **384** (1996) 353.

[2] T. Sjöstrand, *Comp. Phys. Comm.* **67** (1994) 74.

[3] D. Decamp et al. (ALEPH Collab.), *Nucl. Inst. Meth.* **A294** (1990) 121.

[4] D. Buskulic et al. (ALEPH Collab.), *Nucl. Inst. Meth.* **A360** (1995) 481.

[5] The JADE Collaboration, *Z. Phys. C* **33** (1986) 23

[6] D. Buskulic et al. (ALEPH Collab.), *Nucl. Instr. Meth. A* **346** (1994) 461.

[7] D. Brown and M. Frank, *Phys. Lett. B* **313** (1993) 535.