ALEPH 94-004
PHYSIC 94-004
S. D. Johnson
20.01.1994

A Comparison of B-Tagging Algorithms

Sarah Durston Johnson
Niels Bohr Institute - Copenhagen

January 19, 1994

1 Introduction

There are two main algorithms in the ALEPH software to tag b hadrons using lifetime information. These are QIPBTAG [1] and QVSRCH [2]. Before these algorithms were developed, most of the b physics at ALEPH was done by exploiting the semi-leptonic decays of b hadrons and tagging b's via the P_t spectrum of the decay leptons. QIPBTAG uses the larger impact parameters of b hadron decay products to distinguish b hadrons from other hadrons. QVSRCH searches for secondary vertices in events. It tags b hadrons by looking for events with significant secondary vertices. This note describes a direct comparison of these two algorithms when applied to the same 1992 Monte Carlo data set.

The variables used to compare the different b-tagging algorithms were:

- The efficiency of the tag, which is defined to be the number of $Z^0 \to b\bar{b}$ events to pass a given cut divided by the total number of $Z^0 \to b\bar{b}$ events in the data set.
- The purity of the tag, which is defined to be the number of $Z^0 \to b\bar{b}$ events to pass a given cut divided by the total number of events to pass that cut.

The tagging algorithms were compared by examining plots of purity versus efficiency so that one could determine which algorithm gave a better b-purity for a chosen tag efficiency. These variables were examined both for events and for individual hemispheres within events.

2 QIPBTAG

QIPBTAG identifies b hadrons by looking at the signed impact parameter of all tracks produced in a hadronic Z^0 decay. This technique has several advantages over explicitly reconstructing a secondary vertex. First of all, it is conceptually simple. Secondly, tracks from the tertiary vertices, that occur when a b hadron decays into a c hadron, contribute to the statistical power of the impact parameter method, but do not necessarily get assigned to an explicitly reconstructed secondary vertex. QIPBTAG also exploits the fact that tracks with negative impact parameters can be used as a control sample, and so the experimental resolution can be directly measured.

The signed impact parameter $\tilde{D}$ is the distance of closest approach between a track and the b production point, "signed" using the jet direction. The b production point is a primary vertex calculated every event. The QIPBTAG algorithm actually works with the statistical significance of the impact parameter $\tilde{D}/\sigma_D$ to take into account the variation of the statistical resolution of $\tilde{D}$. A probability function is then defined that represents the probability that the measured positive $\tilde{D}/\sigma_D$ is consistent with the fit to the negative significance (from tracks with negative impact parameters). This probability function is then used to tag the b hadrons.

In this comparison, the version of QIPBTAG used was the one included in version 116 of ALPHA. First, a calibration run was performed on the data set (using the CALB card) to get an accurate resolution function. The fit parameters obtained were then used in the subsequent analysis. (A later comparison showed that a negligible decrease in performance occurs if the default calibration fit parameters are used.) The jets used were found using the EFLJ card.

3 QVSRCH

QVSRCH uses a non-standard approach to find secondary vertices in events. The position of the secondary vertex and the assignment of tracks to this vertex are accomplished simultaneously, based on a search in coordinate space. A parameter $\Delta\chi^2$ is defined as the difference between the vertex χ^2 when all the tracks are assigned to the primary vertex and the sum of the primary and secondary vertex χ^2 values when some of the tracks are moved to the secondary vertex. QVSRCH calculates $\Delta\chi^2$ for a grid in secondary vertex coordinate space and calls the point of maximum $\Delta\chi^2$ the secondary vertex. B decays, which generally have significant secondary vertices, will have large $\Delta\chi^2$ values, therefore this parameter can be used to tag b hadrons.

The QVSRCH version used was the most recently available one in the UPHY directory dated June 16, 1993. The beam position and size input into QVSRCH were obtained from QVTXBP and QVTSBP respectively. QVSRCH was then allowed to find its own primary vertex. The jets were found by QVSRCH using pre-clustered jets with the EFLJ card.

4 Comparison of Routines

The comparison was made using 60,000 hadronic 1992 Monte Carlo events generated using JULIA 271 and GALEPH 255 [1]. The efficiency and purity were examined for each tag for individual events and individual hemispheres in an event (defined by the leading jet). QIPBTAG provides both a hemisphere and an event tag variable. In QVSRCH, the event tag variable is the sum of the $\Delta\chi^2$'s for each hemisphere. The event selection cuts

---

[1] A test production produced in November 1993.

required that the routine had a good return value (IRET=0 for QIPBTAG, JERR=0 for QVSRCH) and that the jet (or both jets in the case of the event tag) had a $|\cos\theta| < 0.8$, where θ is the angle the jet makes with the beam axis. For events, these cuts give an overall acceptance of about 69% for both routines.

4.1 Efficiency Vs Purity

Figure 1. shows plots of b-purity versus efficiency for events and hemispheres. QIPBTAG provides a 5-8% increase in b-purity of events for efficiencies in the 45-67% range and a 5-10% increase in b-purity of hemispheres for efficiencies in the 30-62% range as compared to QVSRCH. Figure 2. is a plot of hemisphere b-purity versus efficiency for QIPBTAG, QVSRCH and a tagging method based on the P_t spectrum of leptons from the semi-leptonic decays of b hadrons. The last method used the lepton P_t obtained by the routine QLEPSEL contained in the UPHY package HEVLEP [3]. Even at very low efficiencies ($< 10\%$), the lifetime methods provide higher purities than the lepton tag.

4.2 Angular Distributions

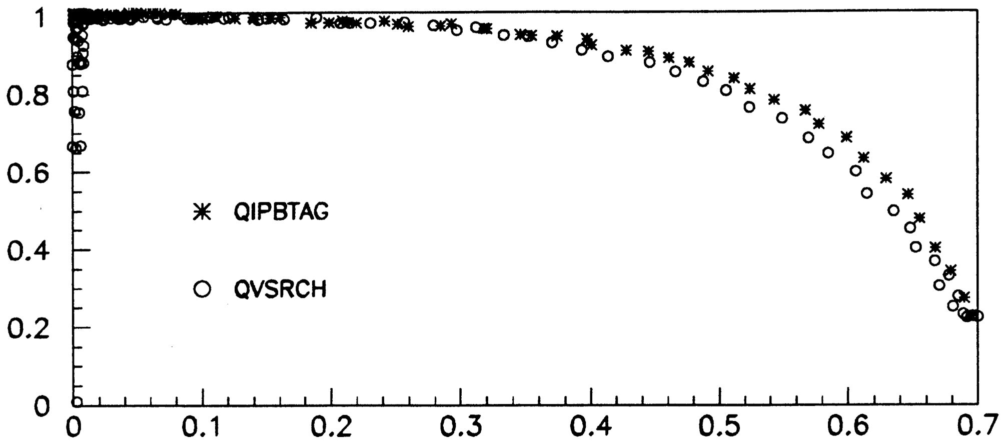
Figure 3. shows plots of hemisphere efficiency and purity versus $\cos\theta$ of the jet for QIPBTAG and QVSRCH. The tag cuts used correspond to an overall efficiency of about 36%. The b-purity is flat with angle for both routines, but QIPBTAG has a lower efficiency than QVSRCH at $|\cos\theta| > 0.7$. This occurs because, unlike QVSRCH, QIPBTAG requires a least one good VDET hit in both the $r - \phi$ and z directions in its track selection cuts and $|\cos\theta| > 0.7$ is near the edge of the VDET acceptance. This seems to indicate that QIPBTAG may be able to improve its efficiency with no loss in purity by loosening the VDET hit requirement.

4.3 Timing

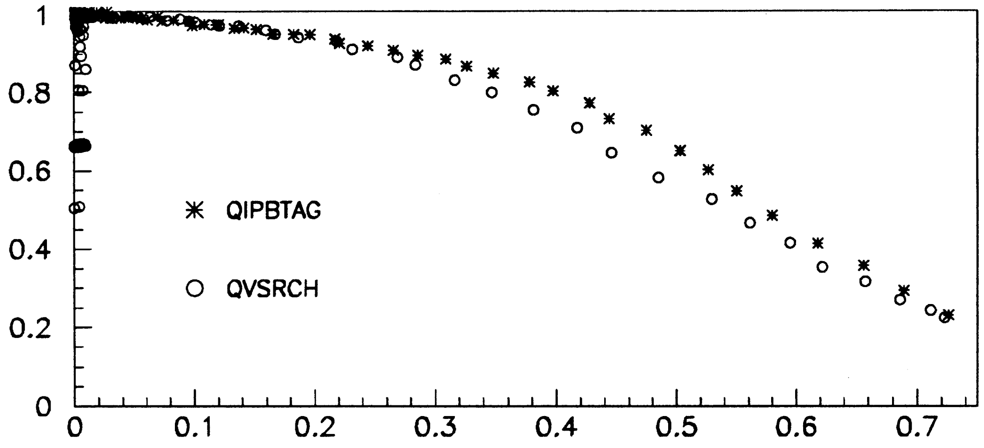
Because QVSRCH explicitly reconstructs two secondary vertices in each event, it takes a longer time to run than does QIPBTAG. On a Vaxstation 4000-90 running on POT format Monte Carlo, QIPBTAG takes 0.031 CPU seconds per event whereas QVSRCH takes 0.111 CPU seconds per event.

5 V^0's in QVSRCH

The version of QVSRCH used does not contain explicit cuts to remove photon conversions, K^0's and Λ^0's. These particles constitute a significant background in $Z^0 \to u\bar{u}, d\bar{d}, s\bar{s}$ events so if one were to remove the tracks from these V^0's a higher b-purity for a given efficiency could possibly be achieved. Figure 4. shows a plot of the difference in QVSRCH event and hemisphere purity, with and without the V^0 tracks included in the secondary vertex calculation, versus efficiency. The V^0 tracks were identified using the QFNDV0 rou-



Event b–purity vs efficiency



Hemisphere b–purity vs efficiency

Figure 1: A comparison of the purity versus efficiency curves for QIPBTAG and QVSRCH, for events and for hemispheres.

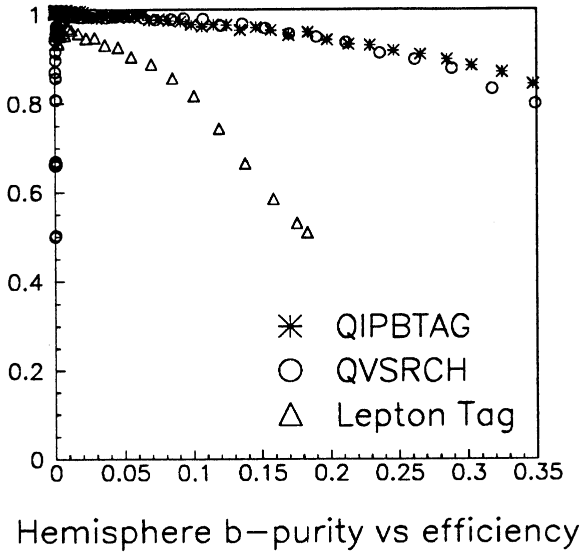


Figure 2: A comparison of lifetime tagging methods with a tagging method based on the P_t spectrum of leptons from semi-leptonic b decays.

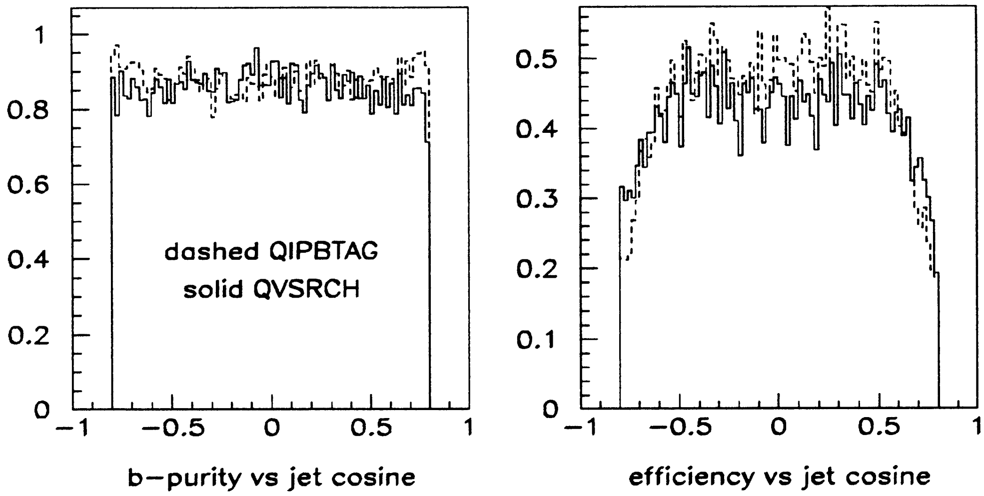


Figure 3: Hemisphere efficiency and purity versus $\cos\theta$ of the jet for QIPBTAG and QVSRCH.

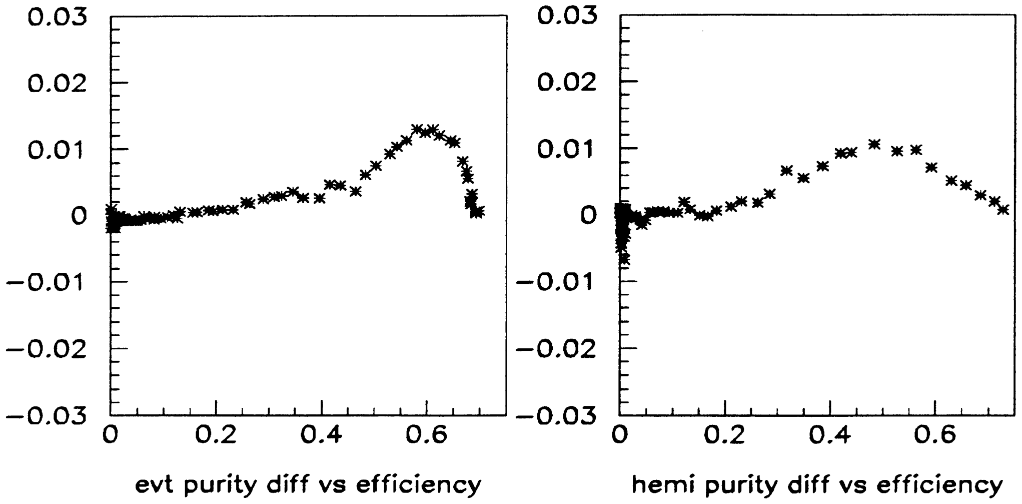


Figure 4: (Purity with V^0 tracks removed - purity without V^0 tracks removed) vs efficiency for QVSRCH for events and for hemispheres.

tine contained within QIPBTAG. Up to a 1% improvement in the b-purity was obtained with the removal of the V^0 tracks.

6 Conclusions

The two major routines to tag b hadrons using lifetime information have been compared. QIPBTAG was seen to perform somewhat better than QVSRCH. QIPBTAG provided a purer b-sample in a certain efficiency range and was comparable to QVSRCH outside this range. QIPBTAG was also seen to run more than 3 times faster. The efficiency of QIPBTAG, however, was seen to have a larger decrease at $|cos\theta| > 0.7$ than QVSRCH, while both routines show a constant purity over the angular range examined. It was seen that QVSRCH's tagging ability can be improved slightly by removing tracks associated with V^0's.

7 Acknowledgements

Thanks to Dave Brown for suggesting this study and his help throughout it.

References

[1] D. Brown and Markus Frank, *Tagging b Hadrons Using Track Impact Parameters*, ALEPH 92-135, PHYSIC 92-124

[2] T. Mattison, *QVSRCH A Tool for Inclusive Secondary Vertex Finding*, ALEPH 92-173, PHYSIC 92-155

[3] P. Perret etal., *Lepton and Jet Definitions for the Lepton Paper*, ALEPH 92-101, PHYSIC 92-90