SW9803

DAPNIA/SPP 97-28

November 1997

COMBINED b -TAGGING

G. Borisov

DAPNIA

Submitted to Nuclear Instruments & Methods in Physics Research A

Le DAPNIA (**D**épartement d'**A**strophysique, de physique des **P**articules, de physique **N**ucléaire et de l'**I**nstrumentation Associée) regroupe les activités du Service d'Astrophysique (SAp), du Département de Physique des Particules Élémentaires (DPhPE) et du Département de Physique Nucléaire (DPhN).

Adresse : DAPNIA, Bâtiment 141
 CEA Saclay
 F - 91191 Gif-sur-Yvette Cedex

Combined b -tagging

G.Borisov¹

CEA Saclay, DSM/DAPNIA, France

Abstract

A method of separating b -quarks from other flavours using combined information is described. The various differences between B -hadrons and other particles are accumulated in a single tagging variable. The method guarantees the optimal tagging if the discriminating variables are independent. The tagging procedure is simple and can easily include any number of discriminating variables. Direct comparison between data and simulation confirms the possibility of using this method in physics measurements.

¹ On leave of absence from Institute for High Energy Physics, Protvino, Russian Federation

1 Introduction

The study of B -hadrons and processes involving production of b -quarks has become an important direction of physics measurements in current and future experiments. In addition to the direct interest in the properties of the b -quark, many new particles, like the Higgs boson or top quark, have at least one b -quark among their decay products. However the production of B -hadrons at e^+e^- or pp colliders is hidden in the large background of light (u, d, s, c) quarks. Thus an efficient technique for selecting ('tagging') events with B -hadrons is required, both to detect the new particles and to study B -hadrons themselves.

The B -hadron is significantly different from all other particles. It has large mass, long lifetime, big decay multiplicity, takes more energy than light hadrons from the initial quark, etc.. However, usually only its long lifetime is used for the tagging. For example, in the OPAL experiment the reconstructed secondary vertices are used as the signature of b -production [1], while in DELPHI the tracks with significant impact parameter, resulting from the large distance between the production and decay points of B -hadrons, are exploited [2]. Recently the large mass of B -hadrons was also included in the tagging procedure, although in a different ways, by the SLD [3] and ALEPH [4] experiments.

This paper describes a method of b -tagging which combines all the above-mentioned differences between B -hadrons and other particles in a single variable. The method has a strong mathematical background which guarantees an optimal combination [5]. Any additional information can be included in the tagging easily, and any such addition will improve the tagging performance provided the new discriminating variables are not correlated with the old ones. Applying this method in the physics measurements of the DELPHI experiment significantly improves the b -tagging efficiency with respect to the lifetime tag used before [6], and opens new possibilities for the study of b -quarks.

2 Tagging Technique

This section explains the optimal method for combining different discriminating variables into a single tagging variable, where "optimal" means that the tagging using the resulting variable gives the best possible background suppression for a given signal selection efficiency. Below we give the prove of this statement, the alternative prove can also be found in [5].

We start our consideration from the case of a single discriminating variable x which has different distributions for the signal and background. We denote the probability density function of this variable for the signal events as $f^S(x)$ and that for background as $f^B(x)$. Suppose first that the ratio $R(x) = f^B(x)/f^S(x)$ decreases monotonically as x increases. If events in some band $[x_1, x_2]$ have been selected, then adding all events with $x > x_2$ can only increase the purity of the sample and the efficiency of the signal selection. Thus in this simplest case the optimal tagging with given efficiency ϵ_0 is just the selection of all events above the threshold $x > x_0$ where x_0 is determined by equation $\epsilon_0 = \int_{x>x_0} f^S(x) dx$.

Let us consider now the general case of one variable when $R(x)$ is an arbitrary function and let us define a new variable y as:

$$y = \frac{f^B(x)}{f^S(x)} \quad (1)$$

The probability density functions of this new variable for background and signal, $g^B(y)$ and $g^S(y)$, are expressed as:

$$g^S(y) = \int \delta\left(y - \frac{f^B(x)}{f^S(x)}\right) f^S(x) dx \quad (2)$$

$$g^B(y) = \int \delta\left(y - \frac{f^B(x)}{f^S(x)}\right) f^B(x) dx \quad (3)$$

Using the properties of the δ -function, equation (3) can be transformed as:

$$g^B(y) = \int \delta\left(y - \frac{f^B(x)}{f^S(x)}\right) y \cdot f^S(x) dx = y \cdot g^S(y) \quad (4)$$

Thus, for the new variable y the ratio of probability density functions $R(y) = g^B(y)/g^S(y)$ is expressed in a very simple way: $R(y) = y$. It changes monotonically with change of y , and we come to the previously considered simplest case. We can conclude that the optimal tagging in the case of an arbitrary single discriminating variable is obtained by using transformation (1) and selecting events with $y < y_0$.

We can easily extend this consideration to the case of any number of discriminating variables $x_1, \dots, x_n$. We define the new variable y as:

$$y = \frac{f^B(x_1, \dots, x_n)}{f^S(x_1, \dots, x_n)} \quad (5)$$

where $f^B(x_1, \dots, x_n)$ and $f^S(x_1, \dots, x_n)$ are the probability density functions for background and signal respectively. The probability density functions $g^B(y)$, $g^S(y)$ are expressed as:

$$g^S(y) = \int \delta\left(y - \frac{f^B(x_1, \dots, x_n)}{f^S(x_1, \dots, x_n)}\right) f^S(x_1, \dots, x_n) dx_1 \cdot \dots \cdot dx_n \quad (6)$$

$$\begin{aligned} g^B(y) &= \int \delta\left(y - \frac{f^B(x_1, \dots, x_n)}{f^S(x_1, \dots, x_n)}\right) f^B(x_1, \dots, x_n) dx_1 \cdot \dots \cdot dx_n \\ &= \int \delta\left(y - \frac{f^B(x_1, \dots, x_n)}{f^S(x_1, \dots, x_n)}\right) y \cdot f^S(x_1, \dots, x_n) dx_1 \cdot \dots \cdot dx_n \\ &= y \cdot g^S(y) \end{aligned} \quad (7)$$

Again the ratio $R(y) \equiv g^B(y)/g^S(y) = y$, so that the optimal tagging is obtained by a simple cut $y < y_0$.

In the case of independent discriminating variables we can write:

$$y = \frac{f^B(x_1, \dots, x_n)}{f^S(x_1, \dots, x_n)} = \prod_{i=1}^n \frac{f_i^B(x_i)}{f_i^S(x_i)} = \prod_{i=1}^n y_i, \quad (8)$$

where $f_i^B(x_i)$, $f_i^S(x_i)$ are probability density functions for background and signal for the variable x_i , and $y_i = f_i^B(x_i)/f_i^S(x_i)$. The optimal tagging in this case is constructed as follows:

- for each discriminating variable the probability density functions for background and signal, $f_i^B(x_i)$ and $f_i^S(x_i)$, are defined;

- the new variables y_i are defined as $y_i(x_i) = f^B(x_i)/f^S(x_i)$;
- the combined tagging variable is defined as $y = \prod_{i=1}^n y_i$;
- the events with $y < y_0$ are tagged as a signal, and the value y_0 can be varied to select the desired tagging purity or efficiency.

The tagging procedure defined in this way is simple, and it allows any number of discriminating variables to be included without additional complication. In practical applications the variables are not independent, so that the n -dimensional definition (5) of y should be used. Such an approach was implemented in [7]. Although it gives optimal tagging, in practice it is difficult to define and use n -dimensional functions for $n > 2$.

Another possibility is to select discriminating variables with a reduced correlation and to apply the tagging method explained above as if the variables were independent. The b -tagging described in this paper is built in this way. It is not absolutely optimal because the variables are not independent. But if the correlations between the discriminating variables are small, it is very close to optimal.

3 Discriminating Variables for b -tagging

The method of b -tagging described here was developed for e^+e^- experiments at the Z^0 pole. In such experiments, two B -hadrons are produced in a Z^0 decay. They have high energy and move in opposite directions, so that the jets containing their decay products can be clearly defined. Their decay points ('secondary vertices') are often separated from the Z^0 decay point ('primary vertex') by large distances and can be identified with the help of a precise track detector.

In this method, the discriminating variables are defined for jets with a reconstructed secondary vertex. This allows the properties specific to B -hadrons to be used to separate their decay products from the particles coming from the fragmentation of the b -quark. In addition, requiring a jet with a reconstructed secondary vertex is itself a good selection, as it removes a significant part of the background. In the case of the DELPHI experiment, 84% of jets with a secondary vertex contain a B -hadron, while the corresponding B -hadron selection efficiency is almost 50%. This can be compared with the initial purity of B -hadrons in hadronic Z^0 decays, which is about 21.6%. The reconstructed secondary vertex is required to contain at least 2 tracks not compatible with the primary vertex and to have $L/\sigma_L > 4$, where L is the distance from the primary to the secondary vertex and σ_L is its error.

The description of the discriminating variables is as follows:

The jet lifetime probability, P_j^+ , is constructed from the positively signed impact parameters of the tracks included in the jet and corresponds to the probability for the hypothesis that they all come from the primary vertex [8, 6]. For jets with B -hadrons, this probability is usually very small because the tracks from the B -decay have significant impact parameters. However, jets with c -quarks also sometimes have low values of P_j^+ because of the nonzero lifetimes of D -mesons. This limits the performance of the lifetime tag. The distribution of $-\log_{10}(P_j^+)$ for different quark flavours is shown in Fig. 1a.

The distribution of **the effective mass of the particles included in the secondary vertex, M_s** , is shown in Fig. 1b. The mass of the secondary vertex in a c -jet

is limited by the mass of the D -meson so that M_s seldom exceeds $1.8 \text{ GeV}/c^2$, while the mass in b -jets extends up to $5 \text{ GeV}/c^2$.

The distribution of the rapidity with respect to the jet direction of tracks included in the secondary vertex, R_s^{tr} , is shown in Fig. 1c. Although a B -hadron has on average a higher energy than a D -meson in a c -jet, the rapidities of particles from B decay are on average less than those from c -quark decay. This is because of the higher B -hadron mass and the larger multiplicity of its decay. The secondary vertices in light quark jets are due mainly to wrongly measured tracks. The wrong measurements occur due to multiple scattering in the detector, interactions in the material, etc., so that the tracks included in the secondary vertices of light quark jets are usually soft and their rapidity distribution is shifted to even lower values.

The distribution of the fraction of the charged energy of the jet included in the secondary vertex, X_s^{ch} , for different quark types is shown in Fig. 1d. In the case of B -hadrons, when almost all particles included in the secondary vertex come from B -decay, the distribution of X_s^{ch} is determined by the fragmentation function $f(b \rightarrow B)$. The same is valid for c -quark jets where the distribution of X_s^{ch} is determined by $f(c \rightarrow D)$, which is softer than $f(b \rightarrow B)$. The energy of the secondary vertex is much less in light quark jets than in b -quark jets, as explained above. As can be seen from Fig. 1d, the distributions of X_s^{ch} for signal (b) and background (u, d, s, c) quarks strongly overlap, so the tagging power of this variable is relatively weak. However, it is retained to demonstrate that, with this combination method, the addition of even weakly discriminating variables improves the tagging performance.

4 Combined b -tagging

Two types of background – jets generated by c -quarks and by u, d, s -quarks – are independent and have different distributions of the discriminating variables. Thus the combined B -hadron tagging variable in a jet with a reconstructed secondary vertex is defined as:

$$y = n_c \cdot \prod \frac{f_i^c(x_i)}{f_i^b(x_i)} + n_q \cdot \prod \frac{f_i^q(x_i)}{f_i^b(x_i)} = n_c \cdot \prod y_i^c + n_q \cdot \prod y_i^q \quad (9)$$

where n_c and n_q are the normalized numbers of c -jets and u, d, s -jets with reconstructed secondary vertices respectively ($n_c + n_q = 1$), and $f_i^q(x_i)$, $f_i^c(x_i)$ and $f_i^b(x_i)$ are the probability density functions of the x_i in uds -, c - and b - quark jets.

The products in (9) run over all tagging variables of the given jet. Because the variable R_s^{tr} is defined for each particle included in the secondary vertex, the corresponding ratio of probabilities for each particle enters in equation (9). For the transformations $y_i^c(x_i) = f_i^c(x_i)/f_i^b(x_i)$ and $y_i^q(x_i) = f_i^q(x_i)/f_i^b(x_i)$, we use smooth functions which are obtained from fits to the ratios of the distributions. The jet is tagged as containing a b -quark if $y < y_0$, and the value y_0 can be varied to select the desired tagging purity or efficiency.

Fig. 2 shows the tagging efficiency versus the purity of the selected sample obtained in the DELPHI experiment for different combinations of the discriminating variables. The starting point for all curves corresponds to the initial efficiency and purity of selection of jets with a secondary vertex. It can be seen that the addition of each new variable improves the tagging performance. As mentioned above, the variable X_s^{ch} is very weak

and can hardly be used for tagging by itself. The efficiency-purity curve for this variable alone is shown in Fig. 2 where it can be compared with the corresponding curve for P_j^+ . However, it can be seen that even the addition of such a weak variable still improves the combined tagging. This is an important advantage of the proposed technique. The overlap of background and signal for the variable R_s^{tr} is also big, as can be seen from Fig. 1. But, due to the large number of secondary tracks, adding R_s^{tr} gives a significant gain in tagging efficiency.

Compared with the simple lifetime tag using only P_j^+ , the combined tagging suppresses the background by more than a factor 3 for a b -tagging efficiency of 30% and by about a factor 6 for a b -tagging efficiency of 20%. A very pure b -sample with a purity exceeding 99.5% can be obtained with the sizeable b -efficiency of 20%. Such a high purity opens new possibilities for the study of B -hadrons.

5 Comparison of Data and Simulation

All distributions for this tagging method are taken from simulation, so that good agreement between data and simulation is important for its application. Differences between data and simulation can arise from incorrect descriptions of measurement effects or of physics processes like hadronization and the decays of heavy particles.

The resolution of the track detector is the main source of measurement effects as it determines the efficiency of reconstructing secondary vertices, the number of tracks included in the secondary vertex, etc. Usually it is possible to tune the description of track resolution in simulation to that observed in data. For the DELPHI experiment the corresponding procedure is described in [9].

Concerning the description of physics processes, a large amount of relevant data has been collected in recent years by the LEP experiments, so that it is now possible to tune the event generators which simulate the hadronization and decays of different quarks with high precision. The most recent study performed by DELPHI is published in [10].

These efforts result in a good agreement between data and simulation in the description of the combined tagging.

The availability of a very pure tagged sample allows the discriminating variable distributions for both signal and background to be extracted from data and compared with those used in the simulation. In Z^0 decays, B -hadrons are produced in pairs, each of them usually being isolated in a separate hemisphere of the event. Therefore a clean and almost undisturbed sample of B hadrons can be obtained in one hemisphere after tagging a B -hadron with high purity in the opposite hemisphere. With a contamination of only about 1%, the discriminating variable distributions in the analysed hemisphere then determine those for B -hadrons. After appropriate normalization, they can then be subtracted from the corresponding distributions for the initial sample of events, containing a large proportion of other quark flavours, to obtain the discriminating variable distributions for the background.

The comparison of these distributions in data and in simulation is shown in Figs. 3-5. Fig. 3 shows the distribution of $-\log_{10} P_j^+$ for background (Fig. 3a) and signal (Fig. 3b). This variable is sensitive to the track resolution, and Fig. 3 confirms that the method used to tune the resolution leads to reasonable agreement between data and simulation. Fig. 4 shows the distributions of track rapidity, which depend on the modeling of physics

processes. Again there is good agreement between data and simulation for both signal and background. Finally, Fig. 5 compares the distributions of the combined tagging variable, $-\log_{10} y$, where y is defined by (9). It shows that the simulation describes the performance of the combined b -tag properly, so that it can be used in different physics applications.

6 Conclusions

A powerful method of separating b -quarks from other flavours has been described. It accumulates various differences between B -hadrons and background in a single variable. It has been shown that this method gives optimal tagging if the discriminating variables are independent. The tagging procedure is simple and can be easily extended to include any additional information. The direct comparison of data and simulation confirms the possibility of using this method in physics measurements.

References

- [1] The OPAL Collaboration, K. Ackerstaff *et al.*, Z. Phys. **C74** (1997) 1.
- [2] DELPHI Collaboration, P. Abreu *et al.*, Z. Phys. **C70** (1996) 531.
- [3] E. Etzion representing the SLD Collaboration, *Measurement of R_b at SLD*, preprint SLAC-PUB-7170, May 1996.
- [4] ALEPH Collaboration, R. Barate *et al.*, Phys. Lett. **B401** (1997) 150.
ALEPH Collaboration, R. Barate *et al.*, Phys. Lett. **B401** (1997) 163.
- [5] T.W. Anderson, An Introduction to multivariate analysis, New York Wiley, 1958.
- [6] G.V. Borisov, *Lifetime Tag of Events with B -hadrons with the DELPHI Detector*, preprint IHEP (Protvino), 94-98 (1994).
- [7] W.J. Murray, *Improved B tagging using impact parameters*, internal DELPHI Note, 95-167.
- [8] ALEPH Collaboration, D. Buskulic *et al.*, Phys. Lett. **B313** (1993) 535.
- [9] G.V. Borisov and C. Mariotti, Nucl. Inst. Meth. **A372** (1996) 181.
- [10] DELPHI Collaboration, P. Abreu *et al.*, Z. Phys. **C73** (1996) 11.

Figure 1: Distributions of the four discriminating variables used in the tagging.

Figure 2: b -tagging efficiency versus purity for the selected sample of jets with reconstructed secondary vertices for different combinations of the discriminating variables.

Figure 3: Distribution of $-\log_{10} P_j^+$ for background (u, d, s, c) jets (a), and b -quark jets (b). The points with errors are from the data, the histogram is the simulation prediction. The contribution of light quark jets is shown as the filled histogram in the upper figure.

Figure 4: Distribution of R_s^{tr} for background (u, d, s, c) jets (a), and b -quark jets (b). The points with errors are from the data, the histogram is the simulation prediction. The contribution of light quark jets is shown as the filled histogram in the upper figure.

Figure 5: Distribution of the combined tagging variable $-\log_{10} y$ for background (u, d, s, c) jets (a), and b -quark jets (b). The points with errors are from the data, the histogram is the simulation prediction. The contribution of light quark jets is shown as the filled histogram in the upper figure.