Индекс 3649

NAME		M.S.
~~J. SKARHA~~	~~3771~~	~~318~~
~~S. CHEVULAEV~~	~~5583~~	~~357~~

ПРЕПРИНТ 94-98, ИФВЭ, 1994

INSTITUTE FOR HIGH ENERGY PHYSICS

IFVE-94-98

Fermilab Library
0 1160 0041612 7

FERMILAB
APR
LIBRARY

IHEP 94-98

G.V.Borisov

Lifetime Tag of Events with B-Hadrons with the DELPHI Detector

Protvino 1994

UDK 539.12 М–24

Abstract

Borisov G.V. Lifetime Tag of Events with B-hadrons with the DELPHI Detector : IHEP Preprint 94-98. – Protvino, 1994. – p. 28, figs. 11, tables 3, refs.: 5

We describe the program for the tagging of events with B-hadrons using their long lifetime. Both the method used for the tagging and the software description are given.

Аннотация

Борисов Г.В. Выделение событий с B-адронами по их времени жизни на установке DELPHI: Препринт ИФВЭ 94-98. – Протвино, 1994. – 28 с., 11 рис., 3 табл., библиогр.: 5.

В работе приводится описание техники выделения событий с B-адронами при помощи их долгого времени жизни. В работе описаны как метод, используемый для мечения таких событий, так и имеющееся программное обеспечение.

© Institute for High Energy Physics, 1994.

Борисов Г.В.
Выделение событий с B-адронами по их времени жизни на установке DELPHI.

Оригинал-макет подготовлен с помощью системы LaTeX.
Редактор Е.Н.Горина. Технический редактор Н.В.Орлова.

Подписано к печати 16.09.1994 г. Формат 60 × 90/16.
Офсетная печать. Печ.л. 1,75. Уч.-изд.л. 2,02. Тираж 240. Заказ 25.
Индекс 3649. ЛР №020498 06.04.1992.

Институт физики высоких энергий, 142284, Протвино Московской обл.

References

[1] D.Brown, *QFNDIP, a primary vertex finder*, ALEPH Note, 92-47.

[2] P.Billoir *et al.*, NIM **A241** (1985),115.

[3] D.Brown, M. Frank, *Tagging B Hadrons Using Track Impact Parameters*, ALEPH Note 92-135.

[4] G.Borisov, C.Mariotti, K.Mönig, *New Measurement of* $\Gamma_{b\bar{b}}/\Gamma_{had}$ *with Lifetime Tag Technique*, DELPHI Note 94-90.

[5] DELPHI collaboration, *Measurement of the* $\Gamma_{b\bar{b}}/\Gamma_{had}$ *Branching Ratio of the* Z^0 *by Double Hemisphere Tagging*, preprint CERN-PPE/94-131.

Received September 15, 1994.

Introduction

The study of events with B hadrons has become one of the most important directions of the explorations at LEP. Many significant results have been obtained here by all of the LEP experiments. The DELPHI detector contains a precise vertex detector which makes it possible to select events with B hadrons using the large track impact parameters of their decay products. This technique of b tagging is implemented in the package AABTAG described in this paper. We suppose to give here the necessary reference information for those who are using this package or are going to use it. In the first section the method of tagging B hadrons is presented, in the second part the available software is described, and in the final part some results of the tagging of B hadrons using the statistics collected in 1992 are shown.

1. Tagging Technique

Hadrons containing b quarks have a long lifetime which, according to the latest measurements, is around 1.55 ps. In Z^0 decays, B hadrons are produced with a large momentum. These two factors, accompanied by the large masses of the B hadrons, cause their decay products to have large impact parameters. This quantity can be used as the only tagging variable for the separation of events with b quarks from the rest of the hadronic background.

The impact parameter in the $R-\phi$ plane is defined as the minimum distance between the track trajectory and the primary interaction point (**primary vertex**) in the plane perpendicular to the beam direction ($R-\phi$ **plane**). **The sign of the impact parameter** is defined as positive if the vector joining the primary vertex and the point of the closest approach of the track lies in the same direction as the jet to which the given track belongs. With such a defini-

tion the tracks from the decays of B hadrons have positive impact parameters, whereas non-zero impact parameters arising from the inaccurate reconstruction of particles' trajectories are equally likely to be positive or negative. Therefore the use of tracks with positive impact parameters increases the tagging efficiency.

The position of the primary vertex is important for the correct measurement of impact parameters and we start the description of the tagging technique with an explanation of the method of primary vertex reconstruction.

1.1. Primary vertex

The primary vertex is reconstructed for every hadronic event using the beamspot position as a constraint. The beamspot size is rather small (e.g. in 1992 it was: $\sigma_x \simeq 100\mu m$, $\sigma_y \simeq 10\mu m$) and its inclusion in the fit significantly improves the accuracy.

The algorithm for the reconstruction of the primary vertex is not completely satisfactory. For light quark events (without the secondary vertices) it gives a reasonable estimate of the position of the primary vertex and its error, but for events with b quarks the accuracy of reconstruction is worse, due to the lower track multiplicity from the primary vertex and to the inclusion in the fit of tracks from the secondary decays. Unfortunately we don't know of a better algorithm in the 2-dimensional $(R-\phi)$ case and if such an algorithm is found the corresponding part of the AABTAG package will be changed.

The main difficulty in the reconstruction of the primary vertex is the selection of tracks for the fit. Tracks with the wrong association of hits in the vertex detector (VD), as well as those from secondary decays of long-lived particles or from interactions in the detector material, may spoil the reconstruction. Some procedure for the removal of such tracks from the fit should be applied. In AABTAG the search for these is done in the same way as in [1]. A similar method was also proposed in [2].

For the primary vertex fit we use the tracks with hits in 2 or 3 layers of the VD. Additionally we select tracks which are close to the beamspot position in terms of δ/σ, where δ is the distance from the track to the beamspot and σ is the corresponding error, which includes both the track error and the beamspot size. The confidence coefficient computed for given δ/σ should be greater than 0.05. There is an additional quality cut which suppresses tracks with the wrong VD hit association. We compute the sum over all N_{VD} hits in the vertex detector associated with a given track, $\Sigma_{tr} \equiv \Sigma(d_i^2/\sigma_{VD}^2)$, where d_i is the distance of the closest approach of the track to the VD hit and σ_{VD} is the accuracy of the vertex detector, and we require that $\Sigma_{tr}/N_{VD} \leq 4$.

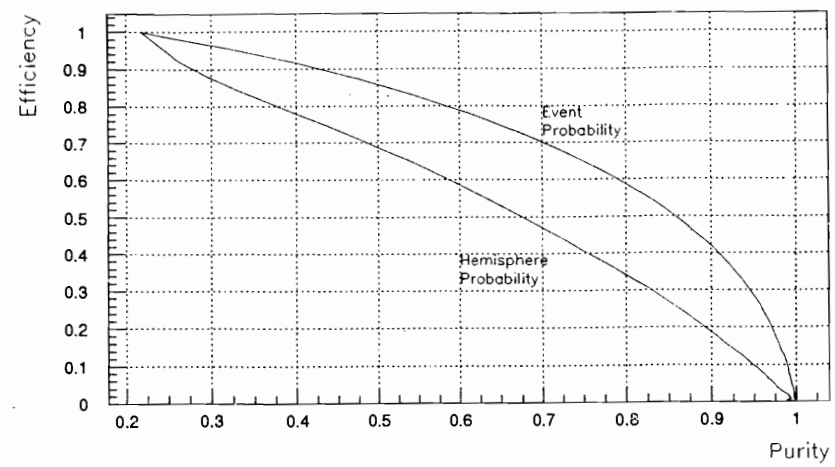


Fig. 10. Tagging efficiency as a function of the purity of the tagged sample

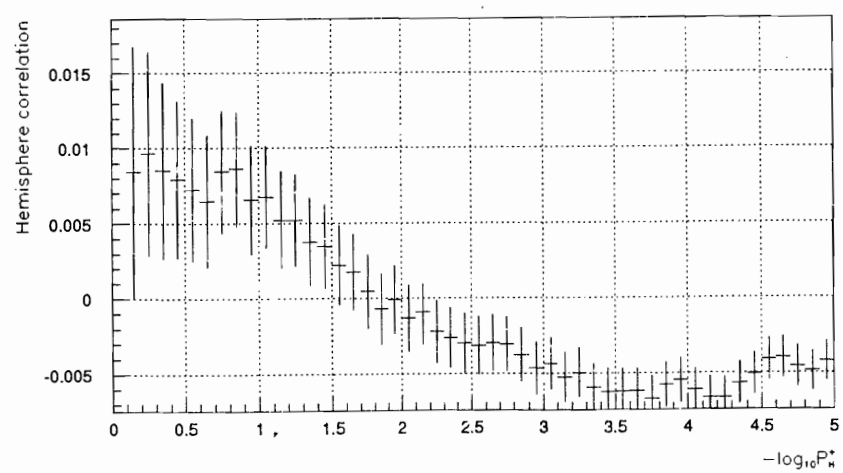
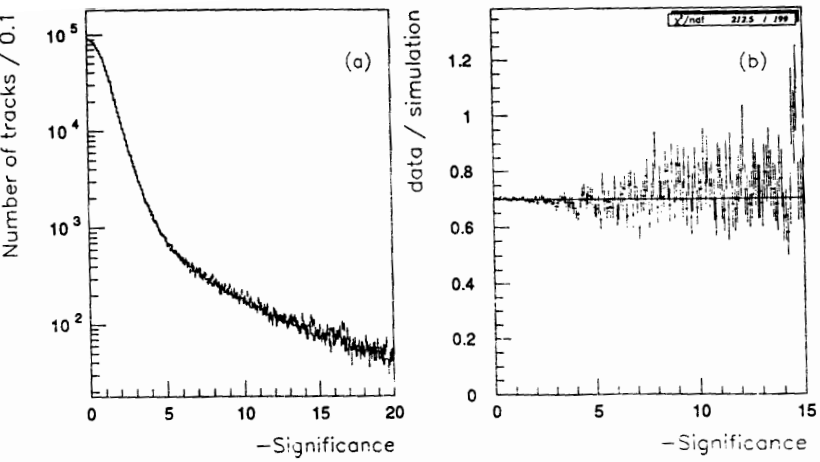
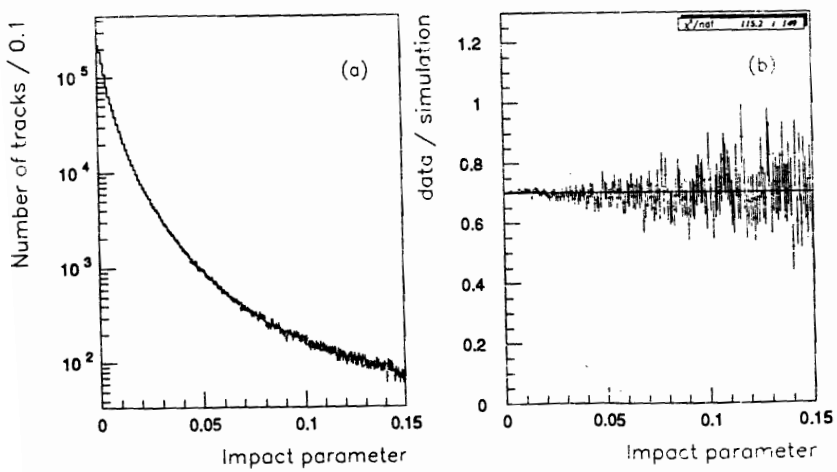


Fig. 11. Hemisphere correlation as a function of tagging cut



. 8. (a): Negative significance distribution for data (points with errors) and for simulation (histogram); (b) The ratio of the significance distributions in data and simulation for tracks with negative impact parameters.



9. (a): Negative impact parameter distribution for data (points with errors) and for simulation (histogram); (b) The ratio of the impact parameter distributions in data and simulation for tracks with negative impact parameters.

At the first step of reconstruction we include in the fit all selected tracks and compute $\chi^2(N_{tr})$ of the fit, where N_{tr} is the number of tracks. After that we exclude each track separately and compute the new $\chi^2(N_{tr}-1)$. We select the track which gives the maximum difference $\chi^2(N_{tr})-\chi^2(N_{tr}-1)$ and exclude it from the fit if the difference exceeds some threshold Δ_{max}. This procedure is repeated while there are tracks for which the difference is greater than Δ_{max}. Because we use the beamspot position as the constraint, all tracks may be rejected from the fit for some events. In this case the primary vertex coincides with the beamspot and the covariance matrix corresponds to the beamspot size. The fraction of such events in a simulated hadronic sample is around 1%.

The primary vertex position is obtained from the minimization of the χ^2 function, which is built similarly to [1]. The distance from the primary vertex to the track (the impact parameter magnitude) in the 2-dimensional case can be written in the form:

$$d_a = \left|\eta_a - (\vec{E}_a \cdot \vec{V})\right|, \tag{1}$$

where (see fig.1) $\vec{E}_a$ is the unit vector from the point $\vec{P}_a$ of the closest approach to the vertex, $\vec{V}$ is the vector from the origin to the primary vertex, and η_a is the projection on to $\vec{E}_a$ of the distance from the origin to the point $\vec{P}_a$ ($\eta_a = (\vec{P}_a \cdot \vec{E}_a)$).

The χ^2 function is written in the form:

$$\chi^2(V_i) = \sum_a \frac{d_a^2}{\sigma_a^2} + \sum_{i=x,y} \frac{(b_i - V_i)^2}{(\sigma_i^b)^2}. \tag{2}$$

In this equation the first sum is taken over all tracks in the fit, d_a is defined by equation (1) and σ_a is the track error, obtained by the error propagation of the (5×5) covariance matrix of track parameters to the point $\vec{P}_a$. The second sum corresponds to the beamspot constraint, where b_i, σ_i^b are the beamspot position and size.

Because the dependence on the vertex position in (2) is quadratic, the minimization of the χ^2 function is very simple. After substitution of (1) in (2) we obtain:

$$\begin{aligned} \chi^2 &= A - 2 \cdot B_i \cdot V_i + C_{ij} \cdot V_i \cdot V_j, \\ A &= \sum_{i=x,y} \left(\frac{b_i}{\sigma_i^b}\right)^2 + \sum_a \left(\frac{\eta_a}{\sigma_a}\right)^2, \\ B_i &= \frac{b_i}{(\sigma_i^b)^2} + \sum_a \frac{\eta_a \cdot (E_a)_i}{(\sigma_a)^2}, \end{aligned} \tag{3}$$

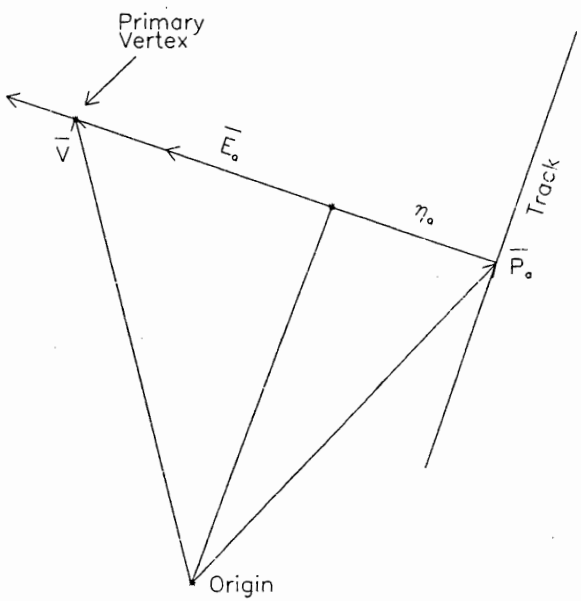


Fig. 1. Definition of impact parameter magnitude.

$$C_{ij} = \frac{\delta_{ij}}{(\sigma_i^b)^2} + \sum_a \frac{(E_a)_i \cdot (E_a)_j}{(\sigma_a)^2}.$$

The values V_i which minimize χ^2 can be easily obtained from (3):

$$\begin{aligned} V_i &= \gamma_{ij} \cdot B_j, \\ \gamma_{ij} &= (C^{-1})_{ij}, \end{aligned} \tag{4}$$

where γ_{ij} is the covariance matrix of the vertex position estimate.

The main advantage of this method of fitting is the quadratic dependence of χ^2 on the fitted values (see (2)). It gives the possibility of analytical and straightforward determination of the vertex position V_i. The simple form of (1) helps also in the calculation of the impact parameter errors (see section 1.2).

Fig.2(a-c) shows the difference between the reconstructed and generated vertex positions in the x-direction for light (uds), charm and bottom quarks obtained with this method. For light quarks the full width at half height of the distribution is 50 μm; for b quarks it is around 90 μm. The same differences divided by the error on V_x are shown in fig.2(d-f). The parameters of the fit of these distributions by the sum of 3 gaussians are given in table 1. For light

Table 3. Hemisphere efficiencies and b purity for tagging with hemisphere probability.

$-\log_{10} P_H^+$ cut	ϵ_{uds}	ϵ_c	ϵ_b	ϵ_b (measured)	b purity
0.5	.286	.484	.781	.764 ± .0494	.398
1.0	.990E-1	.253	.622	.612 ± .0192	.566
1.5	.360E-1	.124	.468	.465 ± .0089	.702
2.0	.132E-1	.553E-1	.333	.334 ± .0047	.806
2.5	.482E-2	.240E-1	.228	.234 ± .0033	.876
3.0	.176E-2	.104E-1	.151	.157 ± .0031	.920
3.5	.672E-3	.436E-2	.0982	.103 ± .0033	.949
4.0	.250E-3	.181E-2	.0626	.0687 ± .0034	.967
4.5	.107E-3	.741E-3	.0386	.0416 ± .0031	.978
5.0	.361E-4	.283E-2	.0235	.0256 ± .0031	.986

tainty in the value of the correlation is around $1.8 \cdot 10^{-3}$ for the tagging cu $-\log_{10} P_H > 2.7$. It is determined mainly by the available simulation statistic and decreases for higher cuts, so that presumably it should not give rise to bi systematic effects in the physical measurements.

Acknowledgements

I would like to thank many people who helped me in the development o this package. Their comments, advice and extensive use of the package signifi cantly improved the original version. I would like to express my acknowledge ments to A.Baroncelli, M.Battaglia, P.Branchini, V.Canale, J.T.M.Chrin, A.D Angelis, G.Gopal, E.Graziani, T.S.Hill, P.Lutz, L.Lyons, M.Margoni, S.Mart M.Mazzucato, F.Simonetto, A.Seitz, T.Spassoff, J.A.Valls Ferrer, S.Überschä W.Venus, M.Witek. The real list of people I am greatly indebted to is muc longer, and my apologies to all those I did not mention above. My speci thanks to C.Mariotti K.Mönig and N.Smirnov, who wrote some parts of tl package, and to A.Normand, who edited this paper. I would also like to thar the management of the DELPHI collaboration and the management of the Se pukhov group of the DELPHI collaboration; without their support this wo would not have been possible.

Table 2. Efficiencies and b purity for tagging with event probability.

$-\log_{10} P_E^+$ cut	ϵ_{uds}	ϵ_c	ϵ_b	b purity
0.5	.324	.598	.916	.399
1.0	.116	.334	.805	.578
1.5	.429E-1	.177	.678	.723
2.0	.162E-1	.880E-1	.549	.827
2.5	.635E-2	.423E-1	.432	.894
3.0	.248E-2	.198E-1	.328	.936
3.5	.959E-3	.930E-2	.244	.961
4.0	.359E-3	.437E-2	.177	.976
4.5	.123E-3	.189E-2	.126	.986
5.0	.582E-4	.808E-3	.088	.991

quarks. For other values of the tagging cut, the efficiencies may be estimated in the same way.

Fig.10 shows the efficiency and purity of the tagged sample for different values of the cut on the event probability (upper curve) and hemisphere probability (lower curve). The efficiency was calculated in simulation for the sample of hadronic events selected within the acceptance of the vertex detector (by a cut on the cosine of the polar angle of the thrust direction: $|\cos(\Theta_{thr})| < 0.75$). This selection corresponds to the efficiency of 69% with respect to the generated sample of hadronic events. Table 2 gives the values of the efficiencies for different flavours and the b purity for different event probability tags, while table 3 gives the variation of the same parameters with the hemisphere probability tag. Additionally, for the hemisphere probability we give the b efficiency measured in the data by the double hemisphere tag method. It is very close to the b efficiency in simulation, which reflects our good understanding of the detector in 1992. The remaining differences may come from physical uncertainties in the production and decay of heavy hadrons.

The technique of many physical analyses could be such that b tagging is performed in one hemisphere, where a pure b sample is selected, and the second hemisphere is used as an unbiased laboratory for the study of B hadrons. For such a technique it is important to have a weak dependence of tagging efficiencies between hemispheres. Fig.11 shows the values of the hemisphere correlation in b events in simulation as a function of the tagging cut on the hemisphere probability. The definition of the hemisphere correlation may be found in [5]. As can be seen, these correlations are small, but some corrections should perhaps be taken into account in the physical analyses. From the studies with double hemisphere tagging [4], we may conclude that the uncer-

Table 1. Parameters of the fit of distributions in Fig.2 by the sum of 3 gaussians. For each gaussian its r.m.s. (σ_i) and fraction (x_i) are given.

Figure	x_1	σ_1	x_2	σ_2	x_3	σ_3
(a)	.356	18.3 μm	.414	42.4 μm	.23	92.3 μm
(b)	.367	23.1 μm	.439	53.1 μm	.194	104.9 μm
(c)	.262	27.8 μm	.425	60.0 μm	.313	112.5 μm
(d)	.375	0.94	.599	1.34	.026	2.79
(e)	.503	1.04	.424	1.61	.073	3.21
(f)	.515	1.10	.346	1.94	.139	4.31

quarks the shape of the distribution is close to gaussian and the difference may be due to the non-gaussian behaviour of the impact parameter distribution (see Fig.3). For b quarks (fig.2f) the distribution has large tails, which are the result of the inclusion in the primary vertex fit of some secondary tracks. As we explained at the beginning of this section, we do not have a better algorithm for suppression of such tracks in 2 dimensions.

1.2. Impact Parameters

The magnitude of the impact parameter is defined by (1). The sign of the impact parameter is defined according to the jet direction. In the package AABTAG there are different options for the jet definition, including the most general possibility of the jet being defined by the user himself. See section 2 for more details. The expression for the impact parameter error may be obtained from (1-4) by the standard procedure of error propagation:

$$\sigma_{d_a}^2 = \sum_c \left(\frac{\partial d_a}{\partial \eta_c}\right)^2 \cdot \sigma_c^2 + \sum_{i=x,y} \left(\frac{\partial d_a}{\partial b_i}\right)^2 \cdot (\sigma_i^b)^2. \tag{5}$$

The partial derivatives in (5) are computed using (3-4):

$$\begin{aligned} \left(\frac{\partial d_a}{\partial \eta_c}\right)^2 &= \left(\delta_{ac} - (E_a)_i \cdot (E_c)_j \cdot \gamma_{ij} \cdot \frac{1}{\sigma_c^2}\right)^2, \\ \left(\frac{\partial d_a}{\partial b_i}\right)^2 &= \left((E_a)_i \gamma_{ij} \frac{1}{(\sigma_i^b)^2}\right)^2. \end{aligned} \tag{6}$$

After substitution of (6) in (5), a very simple result for the impact parameter error can be obtained:

$$\sigma_{d_a}^2 = \begin{cases} \sigma_a^2 - (E_a)_i \cdot (E_a)_j \cdot \gamma_{ij} & \text{if track } a \text{ is included in the fit;} \\ \sigma_a^2 + (E_a)_i \cdot (E_a)_j \cdot \gamma_{ij} & \text{otherwise.} \end{cases} \tag{7}$$

The non-trivial negative sign in the first line of equation (7) arises from the correlation of different tracks in the fitted values V_i.

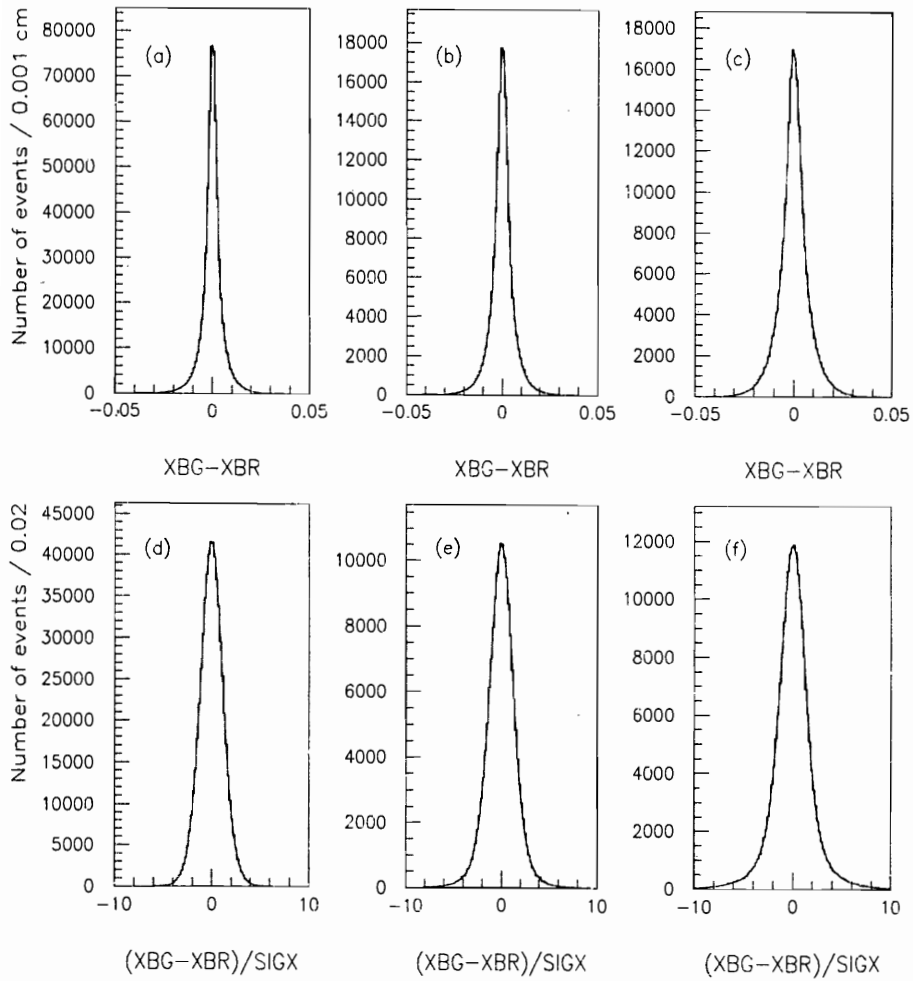


Fig. 2. Difference between the reconstructed and generated vertex positions in the x-direction for (a) light quarks, (b) charm quarks, (c) b quarks; the same difference divided by the error on the reconstructed primary vertex for (d) light quarks, (e) charm quarks, (f) b quarks. The superimposed function is the fit by the sum of 3 gaussians.


```
+CDE, AAPARM.
C     Other COMMONs can be accessed by inclusion of corresponding CDE's
+CDE, AAMNVX.
C ...
C     Definition of variables for event and hemisphere tagging.
      REAL PROBN,PROBP,PROBS,PROBHP(2),PROBHN(2),PROBHS(2),PTHRST(3)
C ...
C     Initialization flag.
      DATA INIT/0/
C ...
C     Initialization phase.
      IF(INIT.EQ.0) THEN
        INIT = 1
        CALL AADATA      ! Initialization of b tagging package.
        IFKOST = .TRUE.  ! Information about reconstructed K0, Lambda
                         ! and e+e- pairs will be taken from ShortDST
        PAR6 = 0.15      ! New cut on the magnitude of impact parameter
      ENDIF
C ...
C     Event and hemisphere probabilities
      CALL BTAGAA(PROBN,PROBP,PROBS)
      CALL AAHEMI(PROBHN,PROBHP,PROBHS,PTHRST)
C     Define the tagging of an event (hemisphere)
      IF(PROBP.LT.0.01) THEN
C ...
C       Work with tagged event.
      ENDIF
C ...
      END
```

3. b tagging with 1992 data

As an example of the use of the b tagging package, we give below some results obtained with statistics collected by DELPHI in 1992. For work with 1992 data we recommend the use of the 92D2 (ShortDST) processing and the tuned (default) version of the AABTAG package. For the tuned version of AABTAG the detector resolution in simulation corresponds to that in data. As an example, Fig.8 shows a comparison of the negative significance distributions for data and simulation, while Fig.9 shows the same comparison of impact parameters. Due to an accurate tuning of the detector resolution, the tagging efficiencies for different flavours can be estimated from the simulation. In [4] the estimate of background efficiencies for light and charm quarks is given for a tagging cut of $-\log_{10} P_H^+ > 2.7$. The uncertainty (which includes the physical uncertainties) is around 7.3% for the light quark efficiency and 9.4% for charm

0.1. The subroutine AACLBR creates and fills calibration histograms for two different sorts of tracks (which are defined by the variable ISRT, see section 2.2). The booked histograms have numbers IHCLBR and IHCLBR+1. The default value of IHCLBR is 2205, but it can be changed after a call to AADATA (+CDE, AAPARM.). At the end of the calibration run the histograms should be written in an HBOOK direct access file (by a call to the HROUT subroutine of HBOOK). The initialization of HBOOK and all operations of saving histograms should be done by the user.

The filling of the calibration histograms should be done separately for the full data set and for at least 50000 simulated events. Fewer statistics are also possible, but the resulting calibration files could be less accurate. After filling the calibration histograms, the user should start a PAW session and execute the MACRO file **CALIB1.KUMAC**, e.g.:

EXEC CALIB1 2205 2206 OUTPUT.DAT

This PAW macro file has 3 parameters: the numbers of the two calibration histograms (2205, 2206) and the name of the resulting calibration file (e.g. OUTPUT.DAT, but may be any other). This calibration file should be assigned to the corresponding LUN instead of the standard one.

2.5. Program Sources

The package AABTAG is included in the DSTANA library. All relevant software is on the BTAGGING disk (CERNVM), or in DISK$DELPHI:[DPADEV. PHY.BTAGGING] (VXCERN). The source PATCHY file is AABTAGXX CARDS (CERNVM), AABTAGXX.CAR (VXCERN) or $DELPHI_PAM/-aabtagxx.car (**shift**). The fortran file may be created using BTAGLIB EXEC (CERNVM) or BTAGLIB.COM (VXCERN). At CERNVM there is a fortran file of the package (AABTAGXX FORTRAN) and a compiled library (AABTAGXX TXTLIB). The macro for making the calibration histograms (as described in section 2.4) is in the file CALIB1.KUMAC. It uses internally the fortran subroutine KUM.FOR. The standard calibration files are on the BTAGGING disk at CERNVM and in the directory DELPHI_DAT: at VXCERN or **shift**.

A simple example of a call to the b tagging package could be the following:

```
      SUBROUTINE USER02
C ...
C     Definition of COMMONs /AAFLAG/ and /AAPARM/ for possible changes
C     of default values of flags and parameters.
+CDE, AAFLAG.
```


The **significance** is defined as the ratio of the impact parameter value to its error. The positive and negative significance distributions are shown in Fig.3. In the ideal case the negative significance distribution should have a gaussian shape. As can be seen from Fig.3, in practice there is a long non-gaussian tail, which is largely due to tracks measured wrongly by the tracking system and partly to particles from the secondary decays and interactions. The gaussian part of the significance distribution is well understood and may be measured directly from the data, while the non-gaussian tail depends significantly on the criteria which are used for the selection of tracks and events. Even such a simple selection as a cut on the thrust axis direction changes the tail. The influence of this change on the tagging results will be discussed later.

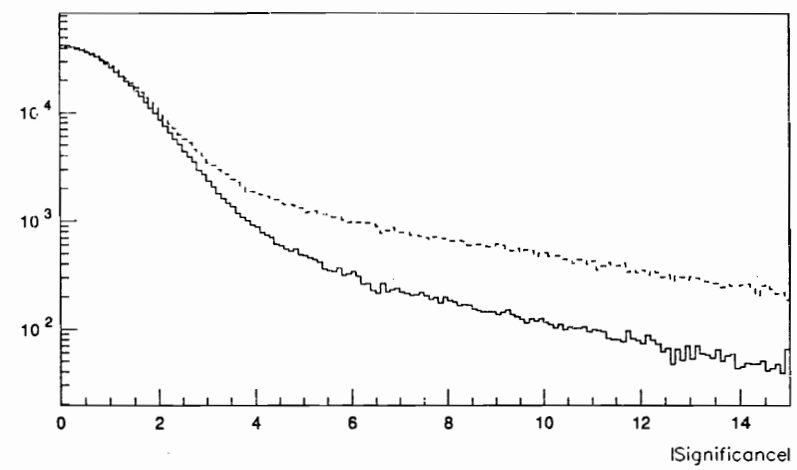


Fig. 3. Distribution of the absolute value of the significance for tracks with negative (solid line) and positive (dashed line) impact parameters for 1991 data.

1.3. Probabilities

For the tagging of B hadrons we use the probability method proposed originally in [3]. It gives the possibility of constructing one tagging variable from all impact parameter values observed in the event. The distribution of this variable for background, which does not contain the decays of long-lived particles, is under control. A very pure sample with a high tagging efficiency can be obtained for events with B hadrons (e.g. 24% tagging efficiency and 96% purity

for events within the acceptance of the vertex detector). The tagging variable is continuous and an optimum choice of the efficiency and purity of the tagged sample can be made for different physical analyses.

For b tagging we select tracks with hits in 2 or 3 layers of the VD, with a cut on the track quality $\Sigma_{tr}/N_{VD} \leq 4$ (the definition of Σ_{tr} was given in section 1.1). The track momentum should be more than 50 MeV (although this cut may be varied). The magnitude of the impact parameter should be less than 0.2 cm. This cut removes particles from the secondary decays of K^0 and Λ and has little effect on the particles from decays of B hadrons.

As previously discussed, particles from the decays of B hadrons should have large positive impact parameters. This fact is demonstrated in Fig.3, which shows the distribution of positive and negative significances. The excess of tracks with positive significance is clearly seen. This tail is used for tagging B hadrons.

The negative significance distribution mainly reflects the detector resolution and is used to build the **track probability function** $P(S_0)$, which is by definition the probability for a track from the primary interaction to have a significance with absolute value S_0 or greater. Mathematically this function is obtained for negative values of the significance with integration of the negative significance distribution from the lowest value up to S_0 and by assuming that for positive significance $P(S_0)$ should be the same:

$$P(S_0) = \begin{cases} \int_{S<S_0} f(S)dS & \text{if } S_0 < 0 \\ P(-S_0) & \text{if } S_0 > 0. \end{cases} \tag{8}$$

Here $f(S)$ is the probability density function of the significance distribution, which is also called the **resolution function**. In practice, the resolution function is obtained from the distribution of tracks with negative impact parameters in a sample of events with a reduced B hadron content. This reduced content is achieved by the anti-b cut $P_E^+ > 0.1$. The definition of the variable P_E^+ and explanation of this cut will be given below in section 1.4. Here we just mention that it reduces the fraction of b events in the sample by more than a factor of 3, to 6.5%. By construction, the distribution of $P(S_0)$ for tracks with positive impact parameters should have a peak at low values. This is shown in Fig.4.

Using the track probability function, the probability for each track in the event can be computed according to the value of the significance. After that, for any group of N tracks (which may be the tracks from the total event, from one hemisphere, from one or several jets) the **N-track probability** is defined as:

ters. This cut removes the majority of the tracks from decays of K^0 and Λ and from interactions in the material of the detector, and only a small fraction of the tracks from decays of B hadrons.

PAR7(=100.) is the cut (in radians) on the maximum angle between track and jet in the $R - \phi$ plane. There is no selection by the default value of the cut, but users can test some other possibilities.

PAR9(=0.050) is the cut on the minimum momentum of the tracks used in the analysis.

PAR10(=0.05) is the cut on the confidence coefficient computed for the tracks w.r.t. the beamspot which is used for the selection of tracks for the primary vertex fit (see sec.1.1).

2.4. External Files

In order to work, the package AABTAG needs two **calibration files** which contain tabulated track probability functions for data and simulation. The standard calibration files are assigned automatically to logical unit number (LUN) ILPRBR (=48) for real data and to LUN ILPRBM (=49) for simulation depending on the processed data set. The default assignment of calibration files may be superseded by the call of the following subroutine **after** the call of initialization subroutine AADATA:

CALL AAOPEN(TYPE,FILE,PLACE,IERR)

Parameters:

I	CHARACTER*2	TYPE	Simulation (='MC') or real data (='RD')
I	CHARACTER	FILE	File name
I	CHARACTER	PLACE	Directory name
O	INTEGER	IERR	Error code

This subroutine opens the calibration file, defined by the user, and returns IERR = 0 in the case of successful completion of OPEN operation and nonzero value of IERR otherwise.

The user can make his own calibration files for special data sets. For this purpose he should make a special calibration run with a call to the subroutine **AACLBR** once per event **after** calculation of all probabilities:

CALL AACLBR(PROBP,PCUT)

In this call PROBP is the event probability as returned by BTAGAA and PCUT is the value of the anti-b cut on the event probability, e.g. PCUT =

Flags

Flags are logical variables which can be switched ON (by assigning .TRUE. to their value) or OFF (=.FALSE.). They are defined in COMMON /AAFLAG/ (**+CDE, AAFLAG.**). In parentheses we give the default values of flags.

Flag **IFTCOR** (=.FALSE.) should normally be switched OFF, but for the 1991 simulation, where the lifetime of B hadrons was 1.2ps, it can be switched ON for better agreement in tagging efficiency between data and simulation. In this case the impact parameters of particles from decays of B hadrons are increased in the simulation to the values corresponding to a 1.6ps average lifetime of B hadrons.

If the flag **IFK0LS** (=.TRUE.) is switched ON, the rejection of K^0, Λ and e^+e^- pairs is performed.

For the sake of compatibility with the previous versions of AABTAG we have the flag **IFOLD** (=.FALSE.). Normally it should be switched OFF.

IFRFIX (=.TRUE.) should be switched ON to run the tuned version of AABTAG, which gives a good agreement between data and simulation.

Flag **IFK0ST** (=.FALSE.) should be switched ON for the DELPHI Short-DST format. In this case the information about reconstructed K^0, Λ and e^+e^- pairs is taken from the ShortDST without spending time on a new search. If this flag is switched ON for Full DST tapes, the standard DELPHI RECV0 code is called and the program works much more slowly. If the user wants the program to work more quickly for full DST tapes, he can switch it OFF. In this case the search for V^0 is done by an internal subroutine of the AABTAG package, which is much faster.

If flag **IFJETS** (=.TRUE.) is switched ON, the search for jets is done inside the AABTAG package and the sign of the impact parameter of a given track is determined with respect to the corresponding jet axis. Otherwise (if IFJETS is switched OFF), the sign is determined with respect to the thrust axis.

Parameters

Parameters are REAL numbers which control the selection of tracks for the analysis. They are defined in the COMMON /AAPARM/ (**+CDE, AAPARM.**). In parentheses we give the default value of each parameter.

PAR1(=4) is the cut on Σ_{tr}/N_{VD} (see sec.1.1) for tracks with hits in 3 layers of VD.

PAR2(=4) is the cut on Σ_{tr}/N_{VD} (see sec.1.1) for tracks with hits in 2 layers of VD.

PAR6(=0.2) is the cut on the maximum magnitude of the impact parame-

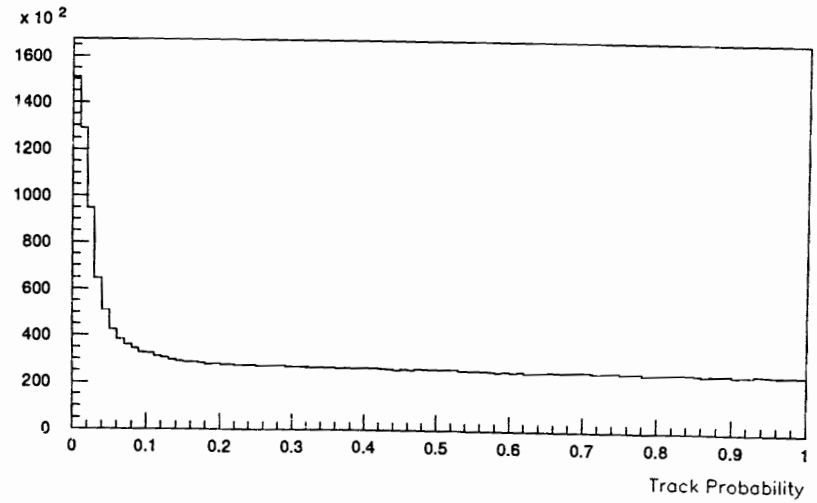


Fig. 4. Distribution of track probability for tracks with positive significance in 1992 data.

$$P_N \equiv \Pi \cdot \sum_{j=0}^{N-1} (-ln\Pi)^j/j!, \text{ where } \Pi \equiv \prod_{i=1}^{N} P(S_i). \tag{9}$$

This variable gives the probability for a group of N tracks with the observed values of significance all to be from the primary vertex. The proof that this variable behaves as a probability is given in [3]. By construction, a flat distribution of P_N is expected for a group of tracks from the primary vertex, provided the significances of these tracks are uncorrelated. If the group includes tracks from secondary vertices, the distribution has a peak at low values of P_N. This is illustrated in Fig. 5, where the distributions of P_E^+ for different flavours are shown. P_E^+ is computed using formula (9) for all tracks in the event with positive significance. As can be seen, the distribution of P_E^+ for light quarks is approximately flat, while for b quarks it has a sharp peak at 0. There is some excess of light quark events at low values of probability, which is caused by residual tracks from V^0 decays or interactions in the detector material, but it is significantly lower than the peaks for c and b quarks.

1.4. Tagging

The N-track probability is the only variable which is used in this approach for tagging B hadrons in a given group of N tracks. The physicist may define

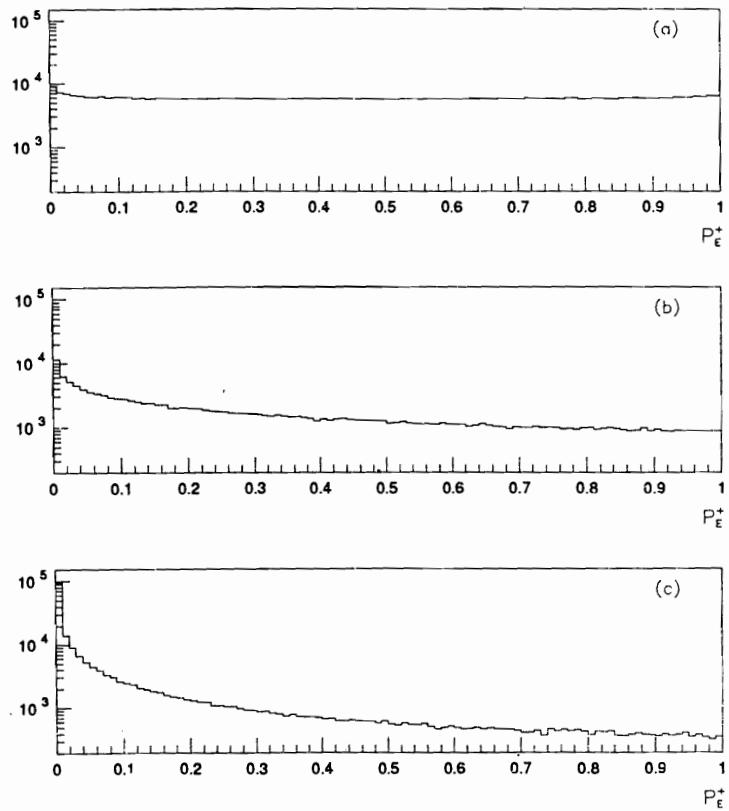


Fig. 5. Event probability for tracks with positive impact parameters for (a) light quark events, (b) charm quark events and (c) b quark events in 1992 simulation.

for himself the group of tracks which he wants to use for tagging. In the AABTAG package there are also the standard tools for computation of the probability using tracks from the total event (**event probability,** P_E) or from two separate hemispheres, defined by the plane perpendicular to the thrust axis (**hemisphere probability,** P_H). The tagging can be done by applying some cut on the N-track probability (e.g. $P_E^+ < 0.01$). By varying this cut the physicist may optimize the purity versus efficiency values for the tagged sample or check the dependence of a physical measurement on the cut. In section 3 we provide some results of b tagging with 1992 data.

An **anti-b cut** can also be achieved by demanding that the positive event

+CDE, AAJETS.

PTHR(1:4)	-	Thrust direction (as returned by LUTHRU)
NJET	-	Number of jets
PJET(1:4,1:10)	-	Jet momenta
IJET(I)	-	Assigned jet number for particle I
THRVAL	-	Thrust value
OBLVAL	-	Oblateness value
ITHR(I)	-	Hemisphere number of particle I

Primary vertex reconstruction
+CDE, AAMNVX.

POSVX(1:3)	-	Position of the primary vertex
COVVX(1:6)	-	Covariance matrix of the primary vertex
CHI2VX	-	χ^2 of the primary vertex fit
NDOFVX	-	Number of degrees of freedom
NATTVX	-	Number of tracks used in the fit
INMVX(I)	-	INMVX(I)=.TRUE. if track was used in the fit
CHI2TR(I)	-	χ^2 for each track

Impact parameters and probabilities
+CDE, AAMAIN.
+CDE, AAMNVX.
+CDE, AAJETS.

PARIMP(I)	-	Impact parameter of the track I
SIGIMP(I)	-	Impact parameter error
PHIV(I)	-	Azimuthal angle between track and associated jet with the sign of the impact parameter (the sign of the impact parameter is positive if PHIV(I) is positive)
TRPR(I)	-	Track probability (defined only for tracks with nonzero ISRT(I))

2.3. Flags and Parameters

The work of the AABTAG package is controlled by a set of parameters and flags. The default values of all parameters were carefully selected. We hope that they are good for the majority of physical applications. It is possible however, to change their values for some special cases. This should be done at the initialization step **after** the call of subroutine AADATA.

Beamspot position
+CDE, AACPRO.

BEAMOY(1:3)	-	Beamspot position (x, y, z);
SIGBE(1:3)	-	Beamspot width in x, y, z;

Tracks selected for b tagging
+CDE, AAMAIN.

For track number I:

IADTR(I)	-	Address of the corresponding PA bank
ISRT(I)	-	Sort of the track (=0, = 1, = 2):
		0 - track is not used in the analysis;
		1 - track with hits in 3 layers of VD;
		2 - track with hits in 2 layers of VD.
NVDP(I)	-	Number of hits in VD
NLAY(I)	-	Number of layers with hits in VD
CHI2VD(I)	-	Σ_{tr}/N_{VD}
NTRK	-	Total number of tracks

Parameters of the tracks
+CDE, AAMAIN.

For track number I:

PAR(1:5,I)	-	Track parameters in the standard DELPHI form (ϵ, z_0, Θ, ϕ, $1/R$)
WGT(1:15,I)	-	Weight matrix of track parameters
TRACK(1:5,I)	-	Internal matrix of track parameters (REAL *8)
TRACK(1,I)	-	x, y-coordinates of the centre of the circle
TRACK(2,I)	-	of track trajectory

Thrust and jets in the event

For the computation of thrust and jets in the event the charged particles and photons are used. The selection of particles is similar to the standard DELPHI selection. The jets are clustered using JETSET 7.3 LUCLUS algorithm with scaled distance $y = m^2/W^2$ used for joining clusters. The distance cutoff is $y_{min} = 0.01$. The hemispheres are defined by the thrust direction and in hemisphere 1 all particles have positive cosine of the angle with the thrust axis.

(hemisphere) probability is greater than some threshold (e.g. $P_E^+ > 0.1$). With such a cut, a sample with a reduced content of B hadrons can be obtained. It can be used, e.g., for the study of background. We use an anti-b cut for building the resolution function, as described above.

Only tracks with positive significance are used for the computation of N-track probabilities for tagging (they are called the **positive N-track probabilities**) because, as discussed earlier, only these tracks contain the lifetime information. All subroutines of the AABTAG package also compute N-track probabilities for tracks with negative significance (**negative N-track probability**). These probabilities may be used for the control of the method. Their distribution is expected to be flat, since tracks with negative significance come mainly from the primary vertex. In practice there are some tracks from secondary decays which have negative significance due to errors in the reconstruction of the primary vertex or jet direction. The remaining correlation between track impact parameters due to mismeasurement of the primary vertex may also be important. These factors spoil the flatness of the negative N-track probability, so its distribution may be used as an indicator of disturbances in the tagging method. Fig.6 shows the negative event probability P_E^- for events which pass the anti-b cut $P_E^+ > 0.1$, explained above, for data (Fig.6a) and simulation (Fig.6b), which should be compared with the negative (Fig.6c) and positive (Fig.5a) probabilities for $Z^0 \to u\bar{u}, d\bar{d}, s\bar{s}$ events. As can be seen, the P_E^- distribution for light quarks is almost flat and coincides with the estimated spectra for both data and simulation. The spike at zero probability due to residual tracks from secondary decays is small, which reflects the quality of the method. The distribution of the positive event probability for light quarks (Fig.5a) is very similar to the negative one (Fig.6c), reflecting the fact that the influence of long-lived particles (K^0, Λ, hyperons) on the N-track probability distribution in $Z^0 \to u\bar{u}, d\bar{d}, s\bar{s}$ events is small.

The resolution functions used for the computation of probabilities are different for data and simulation, but by construction the distribution of the tagging variable is almost flat for the background, which does not contain the decays of long-lived particles. This fact leads to a decrease of the differences between data and simulation in the description of such background. This is one of the main advantages of the proposed method of tagging. It allows one to estimate accurately from the simulation the fraction of background not containing the decays of long-lived particles in the sample of tagged events. The possible uncertainties of this estimate may be obtained from the comparison of tagging efficiency with negative probability in data and in simulation. As can be seen from Fig.6, this efficiency is very close to the true light quark efficiency. The

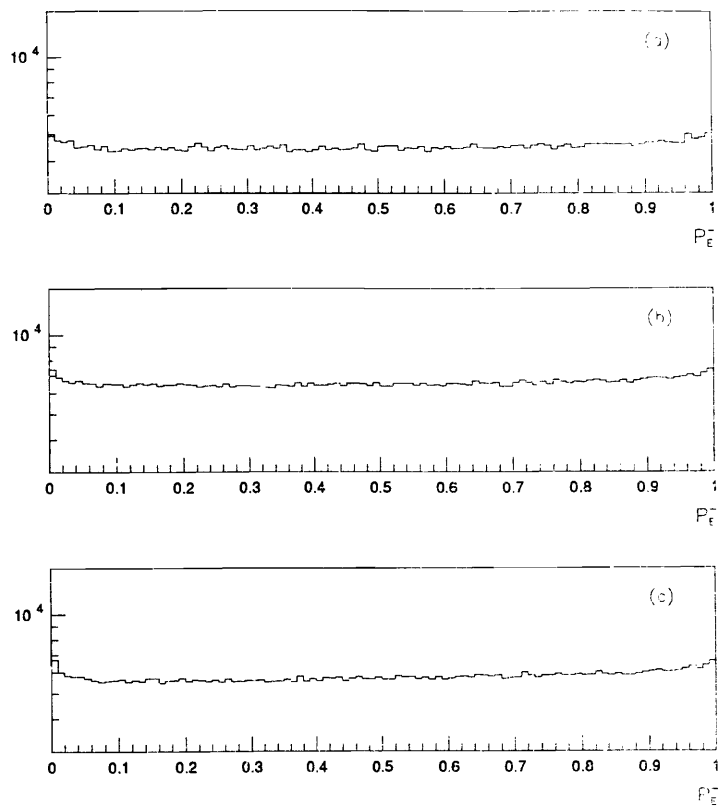


Fig. 6. Negative event probability for (a) 1992 data with cut $P_E^+ > 0.1$, (b) 1992 simulation with cut $P_E^+ > 0.1$ and (c) simulated light quark events.

agreement between data and simulation in the shape of the negative probability is also remarkable. Fig.7a shows the ratio of real data to simulation of the fractions of events selected by the cut on the negative hemisphere probability P_H^- as a function of the cut value C: $-\log_{10} P_H^- > C$. For the most useful values of the cut the difference does not exceed 5%. It is clear from the figure that the simulation may safely be used for the estimate of such background.

The other source of background for the tagging of events with b quarks comes from events containing long-lived particles such as charm mesons. The tagging efficiency for this background depends significantly on the detector resolution. That is why for an accurate estimate of this background a precise tuning of

of the impact parameter is determined by the corresponding jet direction. The jets defined internally in the AABTAG package may also be used. In this case the corresponding information may be found in the COMMON /AAJETS/ described in section 2.2.

Parameters:

I	INTEGER	NTR	Number of tracks selected for the analysis
I	INTEGER	LARR(NTR)	Array of addresses of the track banks (LPA)
I	INTEGER	JETS(NTR)	Jet number to which given track belongs
I	INTEGER	NJ	Number of jets
I	REAL	DJET(3,NJ)	Vectors with jet directions
O	REAL	PRJN	Negative probability
O	REAL	PRJP	Positive probability
O	REAL	PRJS	Probability with all tracks

CALL AANPRB(NTR,LARR,PROBNN,PROBNP,PROBNS)

This subroutine computes the probabilities for any group of tracks. The signs of the track impact parameters are determined with respect to internally defined jets.

Parameters:

I	INTEGER	NTR	Number of tracks selected for the analysis
I	INTEGER	LARR(NTR)	Array of addresses of the track banks (LPA)
O	REAL	PROBNN	Negative N-track probability
O	REAL	PROBNP	Positive N-track probability
O	REAL	PROBNS	N-track probability with all tracks

2.2. Additional variables

All subroutines return the probabilities, but some additional information is available through the COMMONs, defined inside the package with the use of the PATCHY sequences. We give below the description of some useful variables.

Status of the work

+CDE, AAFLAG.

IBAD - Status word.

For some events b tagging cannot be done (no beamspot for given run, primary vertex reconstruction failed, etc.). In this case the variable IBAD has a value different from zero and all returned probabilities are exactly equal to 1.0. Any information from the COMMONs of the b tagging package may be wrong and should not be used.

AABTG1 for every group of tracks, the user should give the number of tracks for the analysis (NTR), the array with the standard LPA addresses of the track banks (LARR), and the beamspot position (XYZ) and size (DXYZ). By defining an artificially large size for the beamspot, the user can effectively remove the beamspot constraint from the fit. The z-coordinate of the beamspot is irrelevant in the b tagging procedure.

Parameters:

I	INTEGER	NTR	Number of tracks selected for the analysis
I	INTEGER	LARR(NTR)	Array of addresses of the track banks (LPA)
I	REAL	XYZ(3)	Beamspot position (x,y,z)
I	REAL	DXYZ(3)	Beamspot size (dx,dy,dz)
O	REAL	PROBN	the same as in BTAGAA
O	REAL	PROBP	the same as in BTAGAA
O	REAL	PROBS	the same as in BTAGAA

Special values of the parameters:

NTR = 0	use all tracks in the event (array LARR is DUMMY);
XYZ(1) = XYZ(2) = 0	get beamspot position from beamspot database;
DXYZ(1) = DXYZ(2) = 0	get beamspot size from beamspot database.

All additional subroutines listed below should be called **after** a call to one of the subroutines BTAGAA, AABTGS, AABTG1.

CALL AAHEMI(PROBHN,PROBHP,PROBHS,PTHRST)

This subroutine computes hemisphere probabilities. Hemispheres are defined by the thrust axis. Its direction is returned in the array PTHRST(3). All tracks in the same direction as the thrust axis are included in the first hemisphere; all tracks opposite to the thrust axis are included in the second hemisphere.

Parameters:

O	REAL	PROBHN(2)	Negative probabilities for two hemispheres
O	REAL	PROBHP(2)	Positive probabilities for two hemispheres
O	REAL	PROBHS(2)	Probabilities with all tracks for two hemispheres
O	REAL	PTHRST(3)	Vector of the thrust direction (normalized to 1)

CALL AAJPRB(NTR,LARR,JETS,NJ,DJET,PRJN,PRJP,PRJS)

This subroutine computes the probabilities for one or several jets. The user provides the number of tracks (NTR), the reference addresses of PA banks (array LARR), the number of jets (NJ), their directions (array DJET, arbitrary normalization) and the association of particles to jets (array JETS). The sign

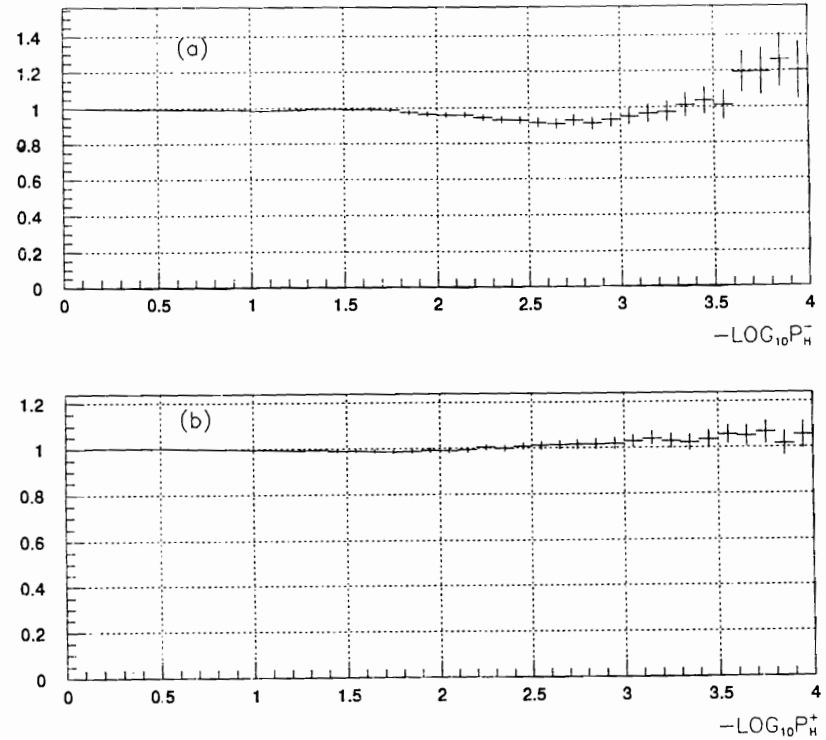


Fig. 7. (a): Ratio of efficiency of tagging by negative hemisphere probabilities in 1992 data to simulation; (b) Ratio of efficiency of tagging by P_H^+ for charm quark in 1992 simulation with resolution taken from data to that with resolution taken from simulation.

the detector resolution in the simulation is needed. A special technique for the tuning of the DELPHI detector resolution was developed. It will be described in a future publication. The uncertainty in the estimate of the lifetime background may be determined by applying the resolution obtained in the data to the simulated events. Figure 7b shows the ratio of the tagging efficiency with P_H^+ for charm events in simulation with the resolution function obtained in the data to that with the resolution taken from simulation. The difference between these two efficiencies is small, mainly due to the tuning of the detector resolution.

The measured resolution function may be biased by tracks from the secondary decays. This bias is irreducible but, since it is rather small, we can expect that it is reasonably described in simulation and depends weakly on the uncertainties of physical parameters of the production and decays of heavy

hadrons. This hypothesis may be checked by varying the fraction of b events in the sample used for calibration. This may be achieved by changing the value of the anti-b cut on the positive event probability. For example, 12%, 6.5%, 3.3% of b events remains in the selected sample after the cut on $P_E^+ > 0.05$, 0.1, 0.5 respectively. The stability of the physical result with the change of b fraction in the resolution function could be a good cross check of the tagging method.

Concluding this section we would like to summarize the main steps of the tagging of B hadrons with the probability method.

For every hadronic event we define the primary vertex and compute the impact parameters with respect to this vertex of all tracks used in the analysis. The sign of the impact parameter is determined according to the track position w.r.t. the jet. The distribution of tracks with negative impact parameters in a sample with reduced b content is used for the computation of the track probability function. This function is then applied for the assignment of track probabilities to all tracks in a given event, according to the observed values of significance. Using tracks with positive impact parameters, the N-track probability is computed for any group of tracks (event, hemisphere, jet etc.). This is the only tagging variable and a sample with the desired purity or efficiency can be selected by applying some cut on it. The negative N-track probability can be used for the control of the method and estimation of the background.

2. AABTAG package

The tagging technique described in the previous section is implemented in the AABTAG package. This is a collection of subroutines, written in FORTRAN-77, with external references to PHDST, VDCLAP, DSTANA and standard CERN libraries only. Below we give a description of the subroutines for common use, some useful COMMONs defined in this package, and different options for its application.

2.1. Subroutines

CALL AADATA

This subroutine should be called once at the beginning of the job for the initialization of the package. The default values of the parameters used in the package can be changed after this call (see section 2.3 below).

CALL BTAGAA(PROBN,PROBP,PROBS)

This is the main subroutine for b tagging. It reconstructs the primary vertex, defines jets and computes event probabilities for tracks with negative (PROBN) and positive (PROBP) impact parameters and for all tracks regardless of the impact parameter sign (PROBS). If there are no tracks with positive or (and) negative impact parameters, the corresponding probability is exactly equal to one. This is a common feature of all subroutines listed below.

Parameters:

O	REAL	PROBN	Negative event probability
O	REAL	PROBP	Positive event probability
O	REAL	PROBS	Event probability for all tracks

CALL AABTGS(XYZ,DXYZ,PROBN,PROBP,PROBS)

This is an alternative way of computating event probabilities. Normally, in BTAGAA the beamspot position is extracted from the beamspot database using the standard tools of the DSTANA library. The subroutine AABTGS allows all computations to be done with an externally defined beamspot. The z-coordinate of the beamspot is irrelevant in the b tagging procedure.

Parameters:

I	REAL	XYZ(3)	Array of beamspot position (x,y,z)
I	REAL	DXYZ(3)	Array of beamspot size (dx,dy,dz)
O	REAL	PROBN	the same as in BTAGAA
O	REAL	PROBP	the same as in BTAGAA
O	REAL	PROBS	the same as in BTAGAA

The subroutines BTAGAA and AABTGS make use of all tracks and, particularly, compute the primary vertex using all tracks selected by the criteria mentioned in section 1.1. Reconstruction of the primary vertex separately for different groups of tracks may be needed for some specific physical analyses. For example, to reduce the hemisphere correlation the reconstruction of the primary vertex may need to be done independently in two hemispheres. This task can be carried out with the help of the following 2 subroutines:

CALL AABTG0
CALL AABTG1(NTR,LARR,XYZ,DXYZ,PROBN,PROBP,PROBS)

These two subroutines should be called instead of BTAGAA or AABTGS. Subroutine AABTG0 should be called once per event. After that, subroutine AABTG1 can be called many times for different groups of tracks. In calling