EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

New Methods of Evaluation of the Flavour Composition in e^+e^- Annihilation by Double Hemisphere Tagging at LEP/SLC Energies

P. Billoir, Ch. de la Vaissière

LPNHE, Laboratoire de Physique Nucleaire des Hautes Energies, Universités Paris VI et VII, 4, Place Jussieu, Tour 33 - RDC, BP 200, 75252 Paris cedex 05, France.

V. Castillo, E. Cortina, E. Higón, F. Martínez-Vidal, J. Salt [1]

IFIC, Centro Mixto Univ. of Valencia − CSIC, Departamento de Física Atómica, Nuclear y Molecular, Avda. Dr. Moliner 50, E−46100 Burjassot, Valencia, Spain.

Abstract

Two new methods are proposed to extract the flavour contents of the events produced at LEP/SLC, together with the classification matrix of a tagging by hemispheres. By utilising the tagging obtained in both hemispheres, the efficiencies, backgrounds and flavour compositions are directly obtained by fitting the data. A minimal dependence on modelling and a consistent treatment of systematic errors are achieved by applying these methods. The choice of the tagging algorithm is irrelevant in the methods, provided that similar efficiencies are reached. As an example, a Multivariate Analysis technique combining the tracking information given by microvertex detector has been applied to extract the $Z \to b\bar{b}$ branching ratio using a standard simulation of a LEP/SLC experiment.

(To be submitted to NIM)

---

[1]Now at CERN supported by a Corresponding Fellow.

1   Introduction

Flavour taggings have become increasingly important in the analysis of hadronic final states produced in e^+e^- colliders at the Z^0 pole and below. In this context, hemisphere tagging appears particularly interesting. The production of the quark-antiquark pair leads to a back to back topology. The tagging algorithm applied to one hemisphere is used to enhance a particular quark flavour, and then the opposite hemisphere is available for studying the properties of that flavour in an essentially unbiassed manner.

The performance of a tagging algorithm is characterized by a set of classification probabilities for each flavour to be classified into several categories. Generally this matrix is taken from Monte Carlo, with the requirement that the simulation should reproduce the physics and the detectors as closely as possible.

This article proposes an alternative. In the case of a tagging applied identically to both hemispheres of the events it is possible to measure from data itself, without explicit reference to simulation, both the flavour composition of the original sample and the classification probabilities. The conditions to be met are:

- Each hemisphere contains the products of one quark or antiquark with the same flavour [2].

- For a given flavour the tagging variables of one hemisphere are not correlated with those in the other.

When these assumptions are almost fulfilled, the previous feature remains true to first order, but simulation is required to evaluate the corrections, coming mainly from residual hemisphere correlations. However, in this case the level of accuracy required for the simulation is not as important as for the usual approach.

Two methods are described in this article and they have been already applied to the measurement of the $Z^0 \to b\bar{b}$ branching ratio in a LEP experiment [1, 2, 3]. In those analyses priority has been put on the measurement of $R_b = \Gamma_{b\bar{b}}/\Gamma_{had}$, but in this description of the methods themselves importance will be also given to the extraction of the classification matrix. The methods will be discussed for an example of classification into three tags. Despite the fact that hadronic events at the Z^0 pole are produced in five flavours, the $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$ flavours have been merged into a single uds family, since the tagging variables used have the same distributions for these flavours. Thus three families of hemispheres will be considered: uds, c and b. Among these families the last one will be privileged because the tagging has been optimized for this sector. The b hemispheres will be mainly distinguished from the others by taking advantage of the long lifetime signature of b hadrons, which require the precise tracking information given by a silicon microvertex detector [4].

The first method is oriented exclusively towards the b sector and it measures the b flavour rate and the classification probabilities of b hemispheres. The second method is more general and extracts also classification probabilities for the other flavours and all tags. The advantage of the knowledge of a complete classification matrix, measured directly from data, is to provide *calibrated* samples of *flavour-enriched* or *flavour-depleted*

---

[2]This condition is not crucial for light quarks (u,d,s). Heavy quarks are almost always produced in opposite directions with respect to the sphericity or thrust axis.

hemispheres without having to rely on simulation to understand the purity and contaminations by other flavours.

In order to test the methods, they have been applied to simulated events of a LEP/SLC experiment. The goal is to measure the flavour composition and the performance of a multivariate tagging algorithm, then to control the quality of this measurement by a comparison with the generated values. The choice of the tagging algorithm is arbitrary, provided the conditions previously mentioned are met. Details about the discriminant variables on which the algorithm is based would be found in appendix A. In section 2 we describe the mathematical formalism of the methods (a glossary of the most relevant mathematical symbols used throughout the text is given in appendix B). The following sections are devoted to the application mentioned above. Section 3 describes the hemisphere definition, the tagging algorithm, its main features and the classification criteria. The results are discussed in section 4. In section 5 we review the main sources of systematic uncertainties. Our conclusions appear in section 6.

2   Mathematical Formalism

Both methods are based on the assumption that the tagging is able to reach high purity for the privileged flavour, i.e. in the b sector. Normally a tagging algorithm is associated with a *classification criterion*, called Δ hereafter and which will be explained in the next section. By imposing a condition on this criterion it is possible to vary the composition and in particular the purity of the tags. We assume that the domain of high purity corresponds to large values of Δ. Such a domain in Δ can be specified by

$$\Delta > \delta \tag{1}$$

where δ is the value of a *purification cut*. If $P_{uds}(\delta)$, $P_c(\delta)$ and $P_b(\delta)$ are the probabilities for the accepted events that a uds, c or b hemisphere is tagged as b for the domain defined by δ, the assumption of 100% purity reached for large values of Δ is equivalent to the following limit conditions:

$$\lim_{\delta \to \infty} \frac{P_{uds}(\delta)}{P_b(\delta)} = 0 \quad , \quad \lim_{\delta \to \infty} \frac{P_c(\delta)}{P_b(\delta)} = 0 \tag{2}$$

2.1   The Asymptote Method

This method intends to measure the fraction R_b as well as the $P_b(\delta)$ tag probability by determining the position of asymptotes in the distributions of several ratios with respect to δ. One critical factor for an accurate evaluation of the asymptotes is to get a sufficient size of the high purity domain.

Let us introduce the fractions of single and double b tagged hemispheres passing the condition (1). An example of these single and double tag fractions is shown in figure 1.a. The fraction of single tags $S_b(\delta)$ may be expressed

$$S_b(\delta) = P_{uds}(\delta)R_{uds} + P_c(\delta)R_c + P_b(\delta)R_b \tag{3}$$

and the fraction $D_{bb}(\delta)$ of double b tagged hemispheres, provided that hemispheres are independent, will be
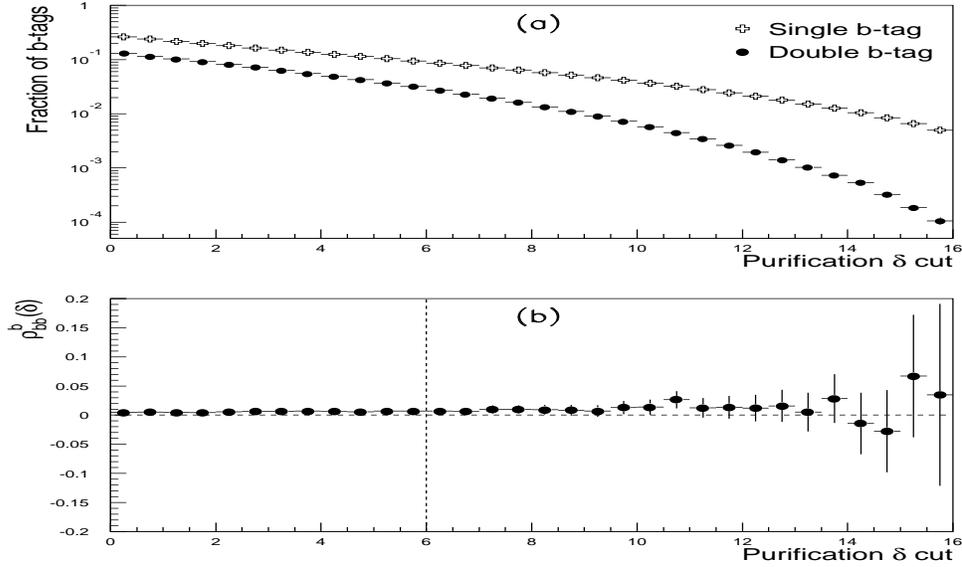
Figure 1: *a) Fractions of single and double b tags $S_b(\delta)$ and $D_{bb}(\delta)$ as a function of the purification δ cut; b) double tag hemisphere correlation factor $\rho_{bb}^b(\delta)$. The value 0.018 ± 0.010 for $\delta = 6.0$ corresponds to the most sensitive correlation coefficient in the matrix method. Note that the distributions are bin to bin correlated.*

$$D_{bb}(\delta) = P_{uds}^2(\delta)R_{uds} + P_c^2(\delta)R_c + P_b^2(\delta)R_b \tag{4}$$

where R_{uds}, R_c and R_b are the flavour fractions in the sample. In order to express these quantities in a more convenient way, we introduce the c and uds probability ratios with respect to b as

$$Q_c(\delta) = \frac{P_c(\delta)}{P_b(\delta)} \quad , \qquad Q_{uds}(\delta) = \frac{P_{uds}(\delta)}{P_b(\delta)} \tag{5}$$

and the flavour fraction ratios $\lambda_c = \frac{R_c}{R_b}$ and $\frac{R_{uds}}{R_b}$. Then we can write

$$S_b(\delta) = P_b(\delta)R_b \left\{ 1 + \lambda_c Q_c(\delta) + \lambda_{uds} Q_{uds}(\delta) \right\} \tag{6}$$

$$D_{bb}(\delta) = P_b^2(\delta)R_b \left\{ 1 + \lambda_c Q_c^2(\delta) + \lambda_{uds} Q_{uds}^2(\delta) \right\} \tag{7}$$

From equations (6) and (7), the ratio

$$r_b(\delta) = S_b^2(\delta)/D_{bb}(\delta) = R_b \frac{\left\{ 1 + \lambda_c Q_c(\delta) + \lambda_{uds} Q_{uds}(\delta) \right\}^2}{1 + \lambda_c Q_c^2(\delta) + \lambda_{uds} Q_{uds}^2(\delta)} \tag{8}$$

tends asymptotically to R_b if the probability ratios $Q_c(\delta)$ and $Q_{uds}(\delta)$ tend towards 0 in the limit of high δ.

The assumption of the independence of hemispheres is measured by the correlation factor $\rho_{bb}^b(\delta)$ defined as

3

$$\rho_{bb}^b(\delta) = \frac{D_{bb}^b(\delta)}{\left\{ S_b^b(\delta) \right\}^2} - 1 \tag{9}$$

where $S_b^b(\delta)$ and $D_{bb}^b(\delta)$ are the single and double fractions computed only for b events. Figure 1.b shows the $\rho_{bb}^b(\delta)$ factor corresponding to the application described in section 3.

2.1.1 Estimation of the $P_b(\delta)$ Probability

If the tag is efficient enough in the b sector, an estimate of the $P_b(\delta)$ probability can be extracted from the data themselves. Let us introduce, instead of $D_{bb}(\delta)$, a more general joint probability to observe $\Delta > \delta'$ in hemisphere 1 and $\Delta > \delta''$ in hemisphere 2,

$$D_{bb}^{joint}(\delta', \delta'') = P_{uds}(\delta')P_{uds}(\delta'')R_{uds} + P_c(\delta')P_c(\delta'')R_c + P_b(\delta')P_b(\delta'')R_b \tag{10}$$

Dividing by equation (6) applied to hemisphere 1, one gets

$$D_{bb}^{joint}(\delta', \delta'')/S_b(\delta') = \frac{P_b(\delta'') \left\{ 1 + \lambda_c Q_c(\delta')Q_c(\delta'') + \lambda_{uds}Q_{uds}(\delta')Q_{uds}(\delta'') \right\}}{1 + \lambda_c Q_c(\delta') + \lambda_{uds}Q_{uds}(\delta')R_{uds}} \tag{11}$$

Let us assume that side 1 is used for tagging, side 2 for counting and δ' is chosen large enough to insure the condition of b purity, i.e. $Q_{uds}(\delta') \approx Q_c(\delta') \approx 0$, then

$$\lim_{\delta' \to \infty} \frac{D_{bb}^{joint}(\delta', \delta'')}{S_b(\delta')} = P_b(\delta'') \tag{12}$$

Once $P_b(\delta)$ is known two new quantities can be constructed

$$s_b(\delta) = \frac{S_b(\delta)}{P_b(\delta)} = R_b \left\{ 1 + \lambda_c Q_c(\delta) + \lambda_{uds}R_{uds}(\delta) \right\} \tag{13}$$

$$d_{bb}(\delta) = \frac{D_{bb}(\delta)}{P_b^2(\delta)} = R_b \left\{ 1 + \lambda_c Q_c^2(\delta) + \lambda_{uds}R_{uds}^2(\delta) \right\} \tag{14}$$

As can be seen their asymptotic limits are also R_b. In particular this limit is rapidly reached in the case of $d_{bb}(\delta)$.

2.1.2 Differential Ratios

Let us consider now a set of values δ_i of the δ cut. The measured observables are $N_b(\delta_i)$ and $N_{bb}(\delta_i)$, which represent the number of single and double tagged hemispheres with $\Delta > \delta_i$. Therefore

$$S_b(\delta_i) = \frac{N_b(\delta_i)}{2N_{tot}} \quad , \qquad D_{bb}(\delta_i) = \frac{N_{bb}(\delta_i)}{N_{tot}} \tag{15}$$

where N_{tot} is the total number of events in the sample. However, adjacent bins are strongly correlated and therefore the ratios $r_b(\delta)$, $s_b(\delta)$ and $d_{bb}(\delta)$ are not bin to bin independent. To avoid this problem uncorrelated ratios tending towards R_b, like the ratios $r_b(\delta)$, $s_b(\delta)$ and $d_{bb}(\delta)$, can be constructed

$$\hat{r}_b(\Delta_i) = \frac{S_b^2(\delta_i) - S_b^2(\delta_{i+1})}{D_{bb}(\delta_i) - D_{bb}(\delta_{i+1})} \tag{16}$$

4

$$\hat{s}_b(\Delta_i) = \frac{S_b(\delta_i) - S_b(\delta_{i+1})}{P_b(\delta_i) - P_b(\delta_{i+1})} \tag{17}$$

$$\hat{d}_{bb}(\Delta_i) = \frac{D_{bb}(\delta_i) - D_{bb}(\delta_{i+1})}{P_b^2(\delta_i) - P_b^2(\delta_{i+1})} \tag{18}$$

with Δ_i defined by $\delta_i < \Delta_i < \delta_{i+1}$. The measurement of R_b is then reduced to the extraction of the asymptotes in the distributions of these ratios. The accuracy on the asymptotic value extraction is determined by statistics but it is also influenced by how rapidly the asymptotic regime is reached, which mainly depends on the tagging efficiency.

2.2 The Matrix Method

This method involves the fit of a matrix of observables. More complex but more general than the asympote method, it is based on the same principles. It measures also the flavour fractions and the hemisphere classification probabilities outside the b sector. The tag probabilities are grouped into a classification matrix C. The objective is to determine the vector R and the matrix C.

The tagging algorithm has the effect to classify the hadronic events into N_T categories. As we shall see below, three flavours require the introduction of at least six categories. Let C_I^l be the *classification matrix* element, *i.e.* the probability to tag a hemisphere of flavour l in the category I ($l = 1, ..., N_F$, where N_F is the number of flavours). The bidimensional array C_I^l is the same for both hemispheres. The same flavour index should be associated to both hemispheres [3].

Applying the tag to both sides and all events, we get a matrix N_{IJ}, number of events classified I and J for hemispheres 1 and 2, shown in figure 2. Dividing by the total number of events one obtains the symmetric matrix of observables D_{IJ} which is the input of the fit discussed in section 4.

If the hemispheres are independent, the expected fraction of events T_{IJ} is written as

$$T_{IJ} = \sum_l C_I^l C_J^l R_l \qquad (I, J = 1, ..., N_T) \tag{19}$$

where R_l is the flavour fraction for a given sample. The aim is to minimize the objective function $G(C, R)$, defined as

$$G(C, R) = \sum_{IJ} \frac{(D_{IJ} - T_{IJ})^2}{\sigma_{IJ}^2} \tag{20}$$

which allow to determine simultaneously the classification matrix C and the R compositions. The fit solution has to be compatible with the constraints: $\sum_l R_l = 1$, $\sum_I C_I^l = 1$, for all l. The matrix of observables itself has to obey the normalization condition $\sum_{IJ} D_{IJ} = 1$. The σ_{IJ} are the statistical errors on D_{IJ} [5].

No solution exists if the number of observables (N_o) is less than the number of unknowns (N_u). For a given N_F and N_T, $N_o = N_T(N_T + 1)/2 - 1$ and $N_u = N_T N_F - 1$. If $N_F = 3$, N_T must be at least 6 (in this case $N_o = 20$, $N_u = 17$).

---

[3]The quark and the antiquark might appear in the same hemisphere when a very hard gluon is radiated. In this case, a tagging based on lifetime does not affect the classification of light flavours.
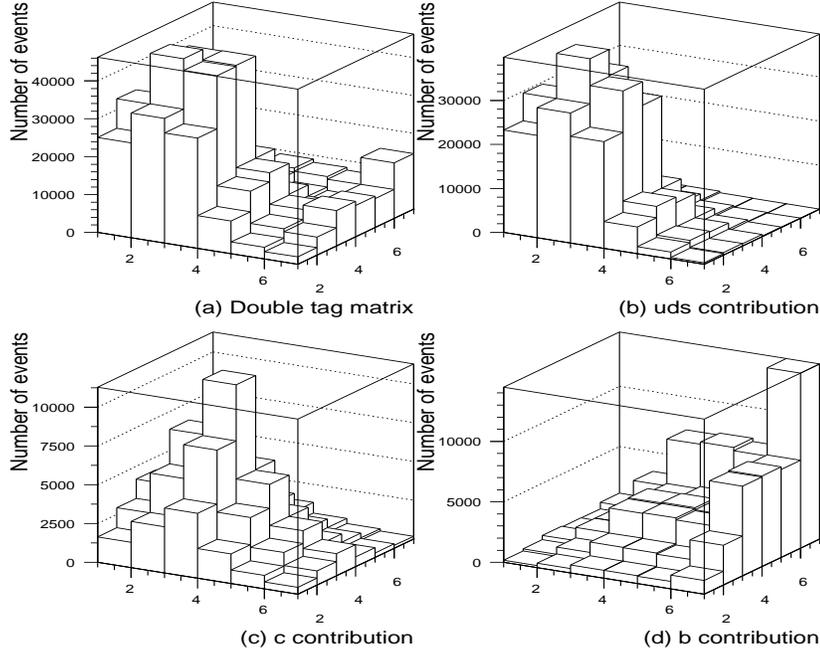
Figure 2: *Population of the double-tag matrix N_{IJ} with their uds, c and b contributions.*

2.2.1 The Rotation Degeneracy

Unfortunately, the minimum of $G(C, R)$ in equation (20) is not unique due to a rotation degeneracy. In fact, if a vector $\vec{V}_I = (C_I^{uds}\sqrt{R_{uds}}, C_I^c\sqrt{R_c}, C_I^b\sqrt{R_b})$ is introduced for each category, all matrix elements can be expressed as a scalar product $T_{IJ} = \vec{V}_I \bullet \vec{V}_J$, which is invariant under rotations in the vector space.

Let us define a vector sum $\vec{U} = \sum_I \vec{V}_I = (\sqrt{R_{uds}}, \sqrt{R_c}, \sqrt{R_b})$ in a 3-D frame, where the three axes correspond to pure *uds*, *c* and *b* states. The vector $\vec{U}$, of unit length, and the set of $\vec{V}_I$ can be viewed as a rigid body. Once a particular solution has been found, other solutions may be generated by moving this rigid body according to three degrees of freedom; two degrees of freedom could be the position (Θ,Ψ dip and azimuth angles) of the extremity of $\vec{U}$ on a sphere of unit radius, the remaining one is an internal rotation γ around the $\vec{U}$ axis. The flavour fractions are then

$$R_{uds} = \cos^2\Theta\cos^2\Psi \;, \quad R_c = \cos^2\Theta\sin^2\Psi \;, \quad R_b = \sin^2\Theta \tag{21}$$

The allowed range of (Θ,Ψ,γ) is limited by two factors. All the C_I^l and R_l elements should be non negative since they are probabilities. Thus, the set of $\vec{V}_I$ vectors should remain in the first octant. When an efficient tagging is reached for a given flavour, some of the $\vec{V}_I$, corresponding to the enriched sample, become practically aligned with a flavour axis. In the limit of three vectors almost aligned with different axes, the rigid body becomes locked. It is observed that the domain of rotations is indeed strongly limited, and the R_l range is actually bound to an interval of a few percent. However the degeneracy does not allow to evaluate the errors.

2.2.2 Resolution of the Degeneracy Problem

An obvious way to solve the degeneracy is to fix N_F parameters which can be taken from simulation, theory or external measurements. The requirement to remain independent of external sources imposes to find other solutions. The most interesting strategy is the following: the degeneracy is broken in the b sector, if some of the C_I^b parameters can be estimated independently (at least 2 in the case of 3 flavours). The third degree of degeneracy can be removed by fixing R_{uds}, R_c or any of the C_I^{uds-c} matrix elements. If X_I^b are estimates of the C_I^b parameters and σ_I their errors, a modified objective function $G_1(C, R)$, including a degeneracy breaking term, can be written as

$$G_1(C, R) = G(C, R) \; + \; \sum_I \frac{(C_I^b - X_I^b)^2}{\sigma_I^2} \tag{22}$$

where the I index only runs over the considered X_I^b.

Let us introduce the relation between the C_I^b and the degeneracy parameters Θ and γ. The structure of the rigid body can be expressed from a particular solution of the fit of equation (20), defined by a vector sum $\vec{U}_0$ and the associated set of vectors $\vec{V}_{0,I}$. In a local frame where the vector $\vec{U}_0$ is chosen as the z axis each vector category $\vec{V}_{0,I}$ is given by three cylindrical coordinates: the projection along the local z axis $g_{0,I} = \vec{U}_0 \bullet \vec{V}_{0,I}$, the radial coordinate $h_{0,I} = \sqrt{\vec{V}_{0,I}^2 - g_{0,I}^2}$ and a local azimuthal angle $\gamma_{0,I}$. For the general solution $\vec{U}(\Theta, \Psi, \gamma)$ the b elements of the classification C_I^b take the simple form

$$C_I^b = g_{0,I} + h_{0,I} \frac{\sin(\gamma + \gamma_{0,I})}{\tan \Theta} \tag{23}$$

The relation between the degrees of freedom Θ and γ is illustrated in figure 3.a, when estimates of the C_I^b are introduced in the left-hand side of these equations. In this figure, where the flavour fraction $R_b(\Theta) = \sin^2 \Theta$ is plotted with respect to γ, the exact solution $\vec{U}_0$ and the generation values C_I^b have been used in equations (23). All category curves intersect at the same point; the third degree of freedom Ψ does not appear in equations (23) and it has no influence on the b sector. As was previously said, it could be removed either by fixing R_{uds} or R_c to their theoretical values or giving to one of the C_I^{uds-c} matrix elements its expected value. In the example that will be described later on, the uds rejection in category 6 is strong and $C_6^{uds} \approx 0$.

The X_I^b estimates are obtained by the technique previously used to calculate the $P_b(\delta)$ probability. From the set of $N_I(\delta)$, which represent the number of hemispheres classified into category I provided that $\Delta > \delta$ on the opposite side, one calculates the fractions

$$f_I(\delta) = \frac{N_I(\delta)}{\sum_J N_J(\delta)} \tag{24}$$

Expressed in terms of classification probabilities $f_I(\delta)$ can be written as

$$f_I(\delta) = \frac{P_{uds}(\delta)C_I^{uds}R_{uds} + P_c(\delta)C_I^c R_c + P_b(\delta)C_I^b R_b}{P_{uds}(\delta)R_{uds} + P_c(\delta)R_c + P_b(\delta)R_b} \tag{25}$$

whose asymptotic limit is

$$\lim_{\delta \to \infty} f_I(\delta) = C_I^b \tag{26}$$

7

The X_I^b are therefore the asymptote values of the $f_I(\delta)$ distributions. In order to avoid the effect of bin to bin correlations, it is better to use the differential ratios

$$\hat{f}_I(\Delta_i) = \frac{N_I(\delta_i) - N_I(\delta_{i+1})}{\sum_J \{N_J(\delta_i) - N_J(\delta_{i+1})\}} \tag{27}$$

which have the same asymptotes as $f_I(\delta)$.
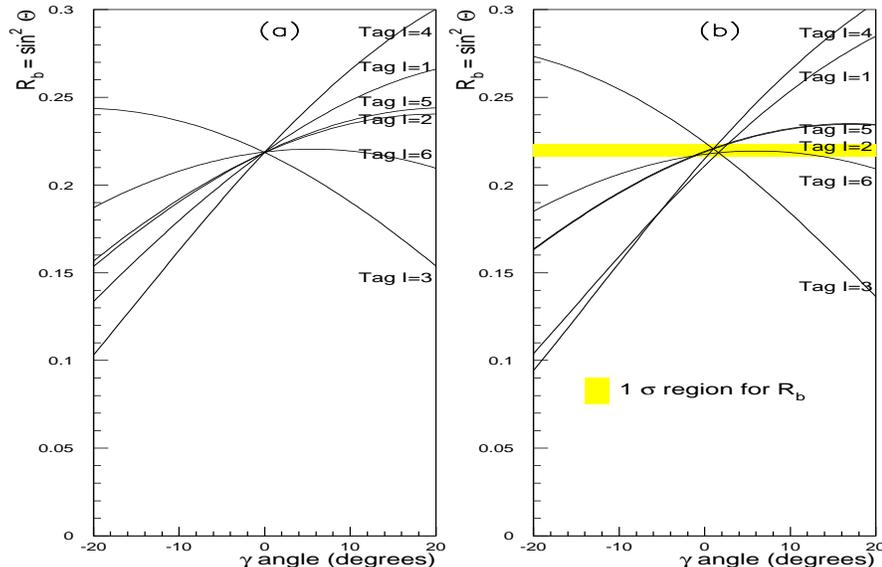
Figure 3: *The R_b fraction as a function of the rotation angle γ of a degenerated solution (Θ, Ψ, γ) for the $G(C, R)$ fit. Each curve is obtained by fixing one of the C_I^b (as labelled on the figure): a) to its true value; b) to the estimated value from the global fit (see section 4.2). The fraction R_b is unaffected by the value of Ψ.*

The figure 3.b summarizes how the degeneracy is solved in the real case to be described in section 4.2. $\vec{U}_0$ and $\vec{V}_{0,I}$ correspond to one solution which minimizes the $G(C, R)$ objective function. Asymptotic X_I^b values of the $\hat{f}_I(\Delta)$ distributions have been given to the left hand side of equation 23 as described.

3   A Tagging Algorithm by Hemispheres

We shall consider now an application to a sample of 540K simulated events after acceptance cuts of a LEP/SLC experiment, supposed to fulfill the requirements mentioned in the introduction [6, 4]. The aim of the acceptance cuts was to discard tracks far away from the interaction region and events outside the microvertex acceptance. The full detector simulated data was generated using [7] with a b lifetime of 1.6 ps. After passing this selection, the composition of the sample was $R_{uds} = 0.6089$, $R_c = 0.1725$ and $R_b = 0.2186$. The impact parameter resolution of the assumed detector is around 20 μm for $\mu^+\mu^-$ pairs.

8

3.1   Hemisphere Definition

Each event was carefully subdivided into two hemispheres as independent as possible. The particles were first distributed into jets using the JETSET standard algorithm (LUCLUS) [8] and the jet direction was given by the jet thrust axis. All particles assigned to jets making an angle of less than 90^0 with the event sphericity axis are attributed to hemisphere one, the rest to hemisphere two. In order to maximize the independence between opposite hemispheres, a primary vertex is computed on each side with an iterative procedure which starts with all the charged particles in the hemisphere. If the fit probability of the global χ^2 is less than 0.05, the particle which contributes with the largest value to the χ^2 is removed, and a new vertex fit is attempted. The process continues until a probability greater than 0.05 is obtained or only two particles remain.

The beam spot position and dimensions have been used as constraints in the vertex fit on both sides, increasing the discriminating power of the tagging but representing a common feature on both hemispheres. As it will be seen in section 5, the inclusion of a beam spot constraint does not seriously correlate the hemispheres.

3.2   The Tagging Algorithm

3.2.1   Description of Variables

We have chosen a multivariate tagging algorithm which has the property of maximizing the available information through a great number of variables. Different processes are expected to lie in different regions of their hyperspace. A set of twelve variables has been defined and evaluated for each hemisphere: one variable (*boosted sphericity*) is computed only with quadrimomenta. Nine variables are based on precise impact parameter information given by a microvertex detector. Two variables mix the two types of information. The definition of the variables and the model distributions are given in appendix A.

3.2.2   Class Likelihoods and Definition of Tags

At the level of a single variable Z, the probability $p^l(z_m)$ to observe a value z in the interval z_m for a hemisphere of flavour l is given by the content $y^l(z_m)$ of the corresponding bin m in a model distribution of this variable and flavour

$$p^l(z_m) = \frac{y^l(z_m)}{N_{tot}^l} \tag{28}$$

where N_{tot}^l is the total number of events in the l flavour distribution. In order to model the shape of these twelve variables we have used a *training sample* of 50K simulated events [9] that was different from the simulated data set used to test the method [4]. Neglecting correlations [5] between individual probabilities, an estimate of the probability to observe the set $\left\{z_{m_1}^1, z_{m_2}^2, ..., z_{m_{12}}^{12}\right\}$, where the superscript corresponds to each variable and the subscript to particular bins, is given for the l flavour by

---

[4] The two samples were produced with slightly different versions and parameter values of the simulation chain.

[5] Correlations reduce the efficiency but not invalidate the classification method.

$$p^l(z_{m_1}^1, z_{m_2}^2, ..., z_{m_{12}}^{12}) = \prod_{\lambda=1}^{12} p^l(z_{m_\lambda}^\lambda) \tag{29}$$

The logarithm of these overall probabilities, called hereafter *class likelihoods* ($L_{uds} = \ln p^{uds}$, $L_c = \ln p^c$ and $L_b = \ln p^b$) are the basis of the classification [10]. The class likelihoods, sorted in decreasing order as L_{first}, L_{second}, L_{third}, are used to tag the hemisphere as uds, c or b according to the highest probability L_{first}.

3.2.3 Classification Criterion and Definition of Categories

In order to define the six categories mentioned in section 2.2 we introduce the classification criterion Δ, defined by

$$\Delta = \ln(p_{first}/p_{second}) = L_{first} - L_{second} \tag{30}$$

Δ being a sensitive indicator of the tag clarity. From the distributions of Δ in the three tags (figure 4), the uds and b tags are subdivided into categories according to the following criteria, while the c has not been subdivided because it is less populated and poorly enriched:

- $uds - tight$: Tag uds and $\Delta > \delta_{uds}$ (category 1)

- $uds - loose$: Tag uds and $\Delta < \delta_{uds}$ (category 2)

- $charm$: Tag c (category 3)

- $b - loose$: Tag b and $\Delta < \delta_b^{low}$ (category 4)

- $b - medium$: Tag b and $\delta_b^{low} < \Delta < \delta_b^{up}$ (category 5)

- $b - tight$: Tag b and $\Delta > \delta_b^{up}$ (category 6)

In order to have similar population in the categories, the chosen values of the cuts are $\delta_{uds} = 2.0$, $\delta_b^{low} = 3.0$ and $\delta_b^{up} = 6.0$.

3.2.4 Main Features

The true values of the R_l flavour compositions and the C matrix elements are shown on the histograms of figure 5. Figure 6 shows the efficiency and contamination of the single and double hemisphere b tags as a function of the purification δ cut. Without any δ cut, the purity for the double hemisphere b tag is already 84% and rapidly approaches 100%. In practice, the D_{66} component of the matrix corresponds to almost pure b events. The following features can be observed:

- uds is the predominant flavour (61% in the analyzed sample). Its purity is over 80% in category 1, but never reaches 100% because the tagging, based essentially on lifetime, is not efficient to discriminate between uds and c flavours at its present level. Therefore, the uds and c columns of the C matrix are similar.

10
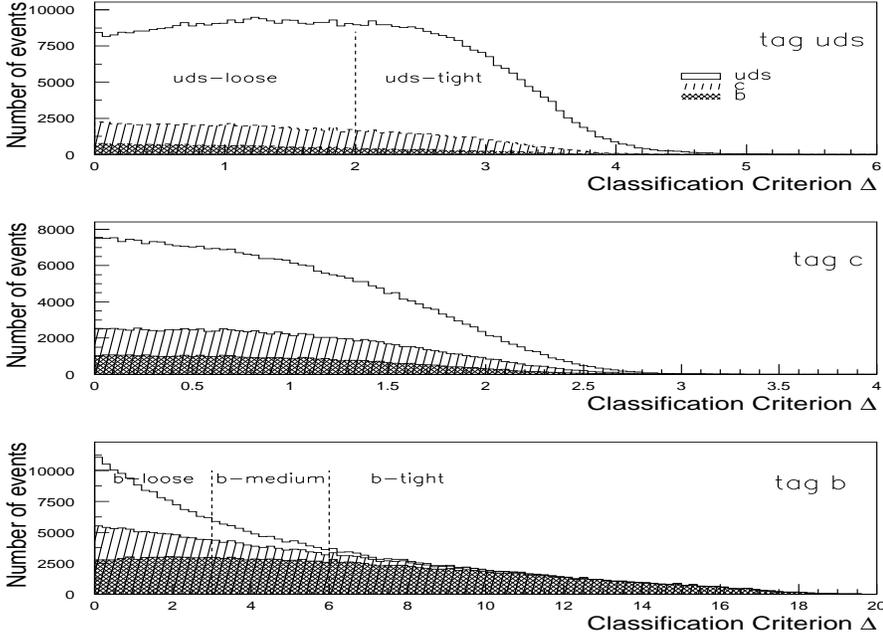
Figure 4: *Distributions of the classification criterion Δ for each tag. Each filled area style shows the different flavour contributions in a given tag. The cut values defining the categories are indicated.*

- The *charm* flavour suffers from being minority (17%) and from overlapping between uds and b in the hyperspace of the twelve tagging variables. A charm hadron travelling enough to be distinguished from uds often appears as a poor b tag. The charm purity reaches 30% at most.

- The *beauty* flavour is also minority (22%), but owing to the decay in cascade $b \rightarrow c \rightarrow s$, the impact parameters and the number of secondary particles are larger. The lifetime tag produces a domain where high purity is achieved.

4 Results of the Methods

4.1 Results of the Asymptote Method

The direct ratio $\hat{r}_b(\Delta)$ is based on the distributions of the single and double b tag fractions shown on figure 1.a. The ratios $\hat{s}_b(\Delta)$ and $\hat{d}_{bb}(\Delta)$ require the estimation of the $P_b(\delta)$ and $\hat{P}_b(\Delta)$ probabilities. In the measurement of R_b, only differential ratios are considered.

4.1.1 Estimates of the probabilities $P_b(\delta)$ and $\hat{P}_b(\Delta)$

The figures 7.a and b show the asymptotic estimation of $P_b(\delta)$ and $\hat{P}_b(\Delta)$. The estimation is based on 22697 hemispheres passing the cut $\delta > 12.0$. The residual contaminations are 1.8% in the c sector and 0.4% in the uds sector for this cut. The quality of the $P_b(\delta)$

11
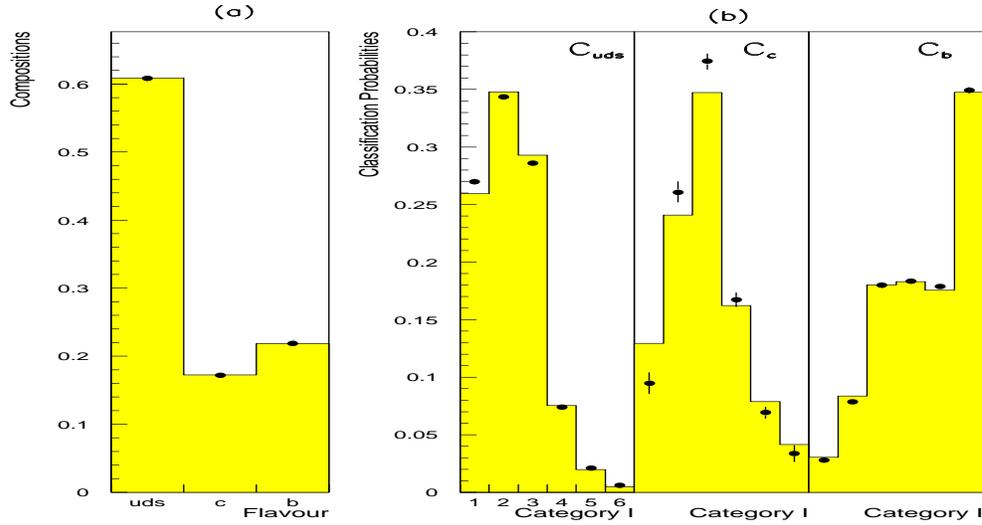
Figure 5: *a) Sample flavour composition; b) elements of the classification matrix. Comparison between generated values (histogram) and fit result (points). A good agreement is observed for the b flavour.*
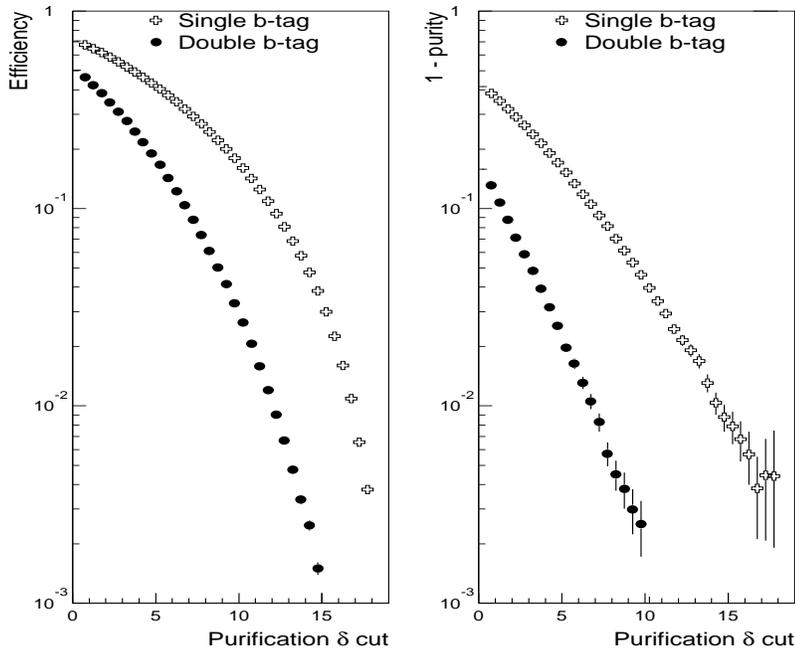
Figure 6: *Efficiency (within the vertex detector acceptance) and contamination of the single and double hemisphere b tags versus the value of the δ cut.*

estimation has been checked by comparing it to its true value, which is the b tag efficiency for b events. Up to a value of $\delta = 14.0$ the measured and true efficiencies agree within 1.5% in relative value (figures 7.c). Similar agreement is observed (figure 7.d) for the differential probability $\hat{P}_b(\Delta)$.
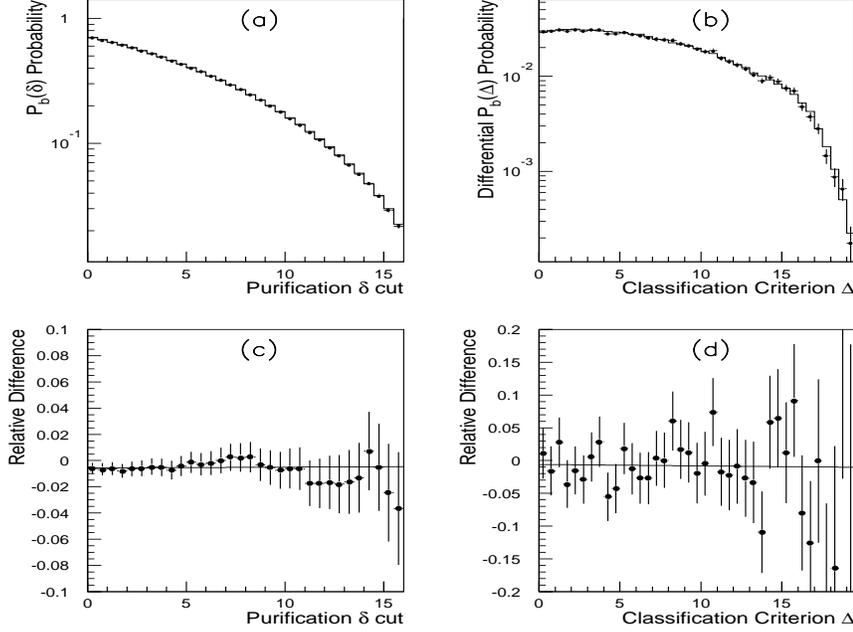
Figure 7: *Asymptotic estimation of the classification probabilities in the b tag; a) $P_b(\delta)$ probability, (solid line represents the expected values); b) differential $\hat{P}_b(\Delta)$ probability; c) relative difference between the estimated and the true integral probabilities; d) same as c) but for differential probabilities.*

4.1.2 Direct Ratio

In figure 8.a, the differential ratio $\hat{r}_b(\Delta)$ defined in equation (16) is plotted versus the classification criterion Δ. The asymptotic behaviour of the curve is visible. An adequate empirical parametrization of $\hat{r}_b(\Delta)$, to extract the asymptote, has been found to be

$$\hat{r}_b(\Delta) = p_0 + p_1 e^{-p_2\Delta - p_3\Delta^2} \tag{31}$$

where p_0 is the asymptotic estimate of R_b and p_1, p_2, p_3 other fit parameters.

The result of such a fit is shown in figure 8.a. The stability of the asymptote has been tested either by varying the Δ range used for the fit or by trying different parametrizations. The estimation of R_b from this ratio is

$$R_b = 0.2185 \pm 0.0052 \pm 0.0023$$

where the first error is purely statistical and the second error is due to the choice of the binning and parametrization. The measurement of R_b can be affected by a correlation

$\rho_{bb}^b(\delta)$ between double and single b tags. This factor, shown in figure 1.b, has been found small in this application and exhibits good stability even at large values of δ. The correction due to the effect of hemisphere correlations is of the order of 1% in relative value and has not been applied here. If more accuracy is required it is possible to rely on Monte-Carlo to correct for this effect. By repeating the procedure for generated b events a specific b ratio can be plotted which should be independent of Δ, and the correction to be applied can be evaluated.
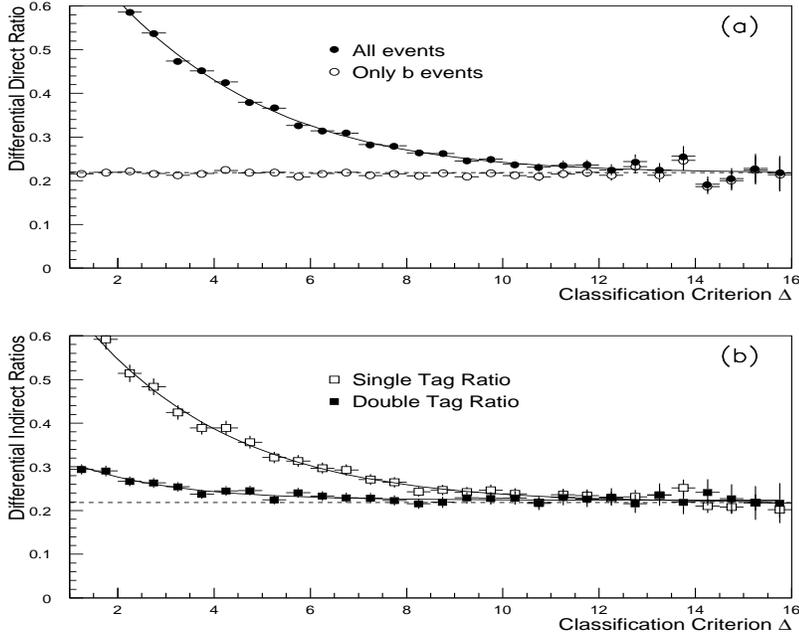
Figure 8: *a) Fit of the differential direct ratio $\hat{r}_b(\Delta)$; b) differential indirect ratios $\hat{s}_b(\Delta)$ and $\hat{d}_{bb}(\Delta)$, defined in equations (17) and (18). The horizontal lines correspond to the expected b fraction.*

4.1.3 Indirect Ratios

The indirect ratios use the estimations of the $P_b(\delta)$ and $\hat{P}_b(\Delta)$ distributions.They exhibit more clearly the asymptotic behaviour. Figure 8.b shows the distributions of the single and double ratios $\hat{s}_b(\Delta)$ and $\hat{d}_{bb}(\Delta)$. The convergence is faster for $\hat{d}_{bb}(\Delta)$ and the asymptote is quite obvious. Both curves appear to be compatible with the same asymptotic value and their dependence with Δ has been parametrized again by a constant plus second order exponential.

The result based on $\hat{s}_b(\Delta)$ is

$$R_b = 0.2162 \pm 0.0060 \pm 0.0030$$

and the evaluation with $\hat{d}_{bb}(\Delta)$ gives

$$R_b = 0.2236 \pm 0.0046 \pm 0.0030$$

14

Both results are consistent with the 0.2186 expected value. Good stability of the asymptote with respect to the fitting range is again observed. The $\hat{s}_b(\Delta)$ and $\hat{d}_{bb}(\Delta)$ happen to be well represented by a constant plus a simple exponential. Taking the average of the two measurements, which are practically independent, one finally quotes

$$R_b = 0.2209 \pm 0.0036 \pm 0.0030$$

4.2 Results of the Matrix Method

The population of the double tagged categories, shown in figure 2.a, is one input of the fit. As can be seen uds and b events populate opposite corners, while the c events overlap largely with uds and b. The set of the $\hat{f}_I(\Delta)$ fractions, plotted on figure 9, is the other input. Good agreement can be seen between the asymptotic limits and the expected C_I^b elements which are also plotted. No other information taken from external sources is introduced in the fit. The assumption that there is no irreducible background from light and c quarks appears verified in the figure 9 (see also figure 4).

4.2.1 Global Fit Procedure

The approach of the $\hat{f}_I(\Delta)$ fractions to their asymptotes can be expressed by the equation

$$\hat{f}_I(\Delta) = C_I^b \; + (C_I^{uds} - C_I^b)\hat{\epsilon}^{uds}(\Delta) + (C_I^c - C_I^b)\hat{\epsilon}^c(\Delta) \tag{32}$$

derived from equation (27), where $\hat{\epsilon}^{uds}(\Delta)$ and $\hat{\epsilon}^c(\Delta)$ are the uds and $charm$ contaminations in the hemisphere used for tagging. In this application, $\hat{\epsilon}^{uds}$ and $\hat{\epsilon}^c$ appear to be well described by second order exponentials of different magnitude but with similar slopes. For that reason the best parametrization of the $\hat{f}_I(\Delta)$ is the exponential like function

$$\hat{f}_I(\Delta) = \; X_I^b \; + \; \frac{a_I}{\sqrt{2\pi}c_I}e^{-b_I\Delta}e^{-\Delta^2/2c_I^2} \tag{33}$$

where the free parameters b_I and c_I give the shape of the distribution function and a_I is a scale factor. Equation (33) can be used to fit separately each of the $\hat{f}_I(\Delta)$ fractions to extract the X_I^b estimates. Then these estimates are introduced in the function $G_1(C, R)$ given by equation (22), which has to be minimized to solve the degeneracy. Figure 3.b suggests how the degeneracy is broken and how category curves intersect when the left hand sides of equation (23) are set to the X_I^b estimates.

The problem with this technique is to evaluate properly the contribution of the systematic errors in the σ_I of the X_I^b which are injected. This difficulty can be avoided if the fits of $\hat{f}_I(\Delta)$ and $G_1(C, R)$ are merged into a single one by minimizing the global objective $G_2(C, R)$ function defined as

$$G_2(C, R) = \sum_{IJ} \frac{\{D_{IJ} - T_{IJ}\}^2}{\sigma_{IJ}^2} \; + \; \sum_{I,\Delta} \frac{\left\{\hat{f}_I(\Delta) - C_I^b - \frac{a_I}{\sqrt{2\pi}c_I}e^{-b_I\Delta - \Delta^2/2c_I^2}\right\}^2}{\sigma_{\hat{f}_I(\Delta)}^2} \tag{34}$$

The classification matrix and the R_l compositions are simultaneously obtained by this way. The $\sigma_{\hat{f}_I(\Delta)}$ is the statistical error on $\hat{f}_I(\Delta)$ for each bin of Δ.

The main advantage of this global fit is to provide the unique solution reproducing simultaneously the tensor population, given by the G function, and the $\hat{f}_I(\Delta)$ fractions appearing in the second term of G_2. Moreover, as it is explained in section 5.1, correlation

15

effects can be simultaneously studied for both terms of equation (34). The disadvantage is the introduction of a relatively large number of auxiliary shape parameters which offer too much flexibility to the fit and increases the statistical errors on the most important unknowns, the R and C elements.

Another point to be commented is the *double counting* of events in the definition of $G_2(C, R)$. Some events, mainly of the b flavour, enter in both terms of equation (34). In order to estimate correctly the statistical error, we have generated data sets by fluctuating randomly the number of events on the cells of the D_{IJ} matrix and on the bins of the $\hat{f}_I(\Delta)$ distributions. The dispersion of the R_b fitted values was taken as the statistical error of the fit. This error agrees within 5% with the estimation given by the $\chi^2 + 1$ method and therefore we conclude that the net effect of this double counting is small.

The function $G_2(C, R)$ has been minimized by fixing the R_c parameter to the generated value of 0.1725. As has already been remarked in section 2.2.2, this constraint has no effect on any parameter of the b sector. The fitted b fraction is found to be

$$R_b = 0.2197 \pm 0.0040$$

in agreement with the generated value and giving $G_{2,min}/d.o.f. = 258.7/225$. The error is only statistical.

Table 1 and figure 5.b compare the fitted values of the classification matrix elements with the expected ones. The first error is statistical and the second one is the systematic due to the R_c uncertainty (calculated with a 8% variation). For the b column of the classification matrix the agreement between the measured and expected values is excellent and of the the same quality than the one observed in the asymptote method for the $P_b(\delta)$ estimation. The agreement is still satisfactory for the *uds* column. The behaviour of the *charm* column is reproduced, but the agreement is only at the 10% level. This lack of accuracy reflects the important overlapping between c and *uds*, the fact that *charm* is minority, and a possible effect of hemisphere correlations.

5   Sources of Systematic Errors

Systematic errors are specific to a given analysis. For this reason this discussion will be limited to an overview of the sources of systematics, in particular the ones relevant for the method. For the same reason, we shall not discuss errors on the classification matrix elements, because these elements are specific to a particular tagging algorithm. The discussion will be limited to the errors made on the R_b fraction and consider only the *matrix method* since it provides its most precise value. The main sources of systematic errors are the hemisphere correlations.

5.1   Effect of Hemisphere Correlations

In order to take into account the hemisphere correlations, the expression of the T_{IJ} fractions in the $G_2(C, R)$ function can be reformulated as

$$T_{IJ} = \sum_l C_I^l C_J^l (1 + \rho_{JI}^l) R_l \tag{35}$$

where

$$\rho_{JI}^l = \frac{D_{IJ}^l}{C_I^l C_J^l} - 1 \tag{36}$$
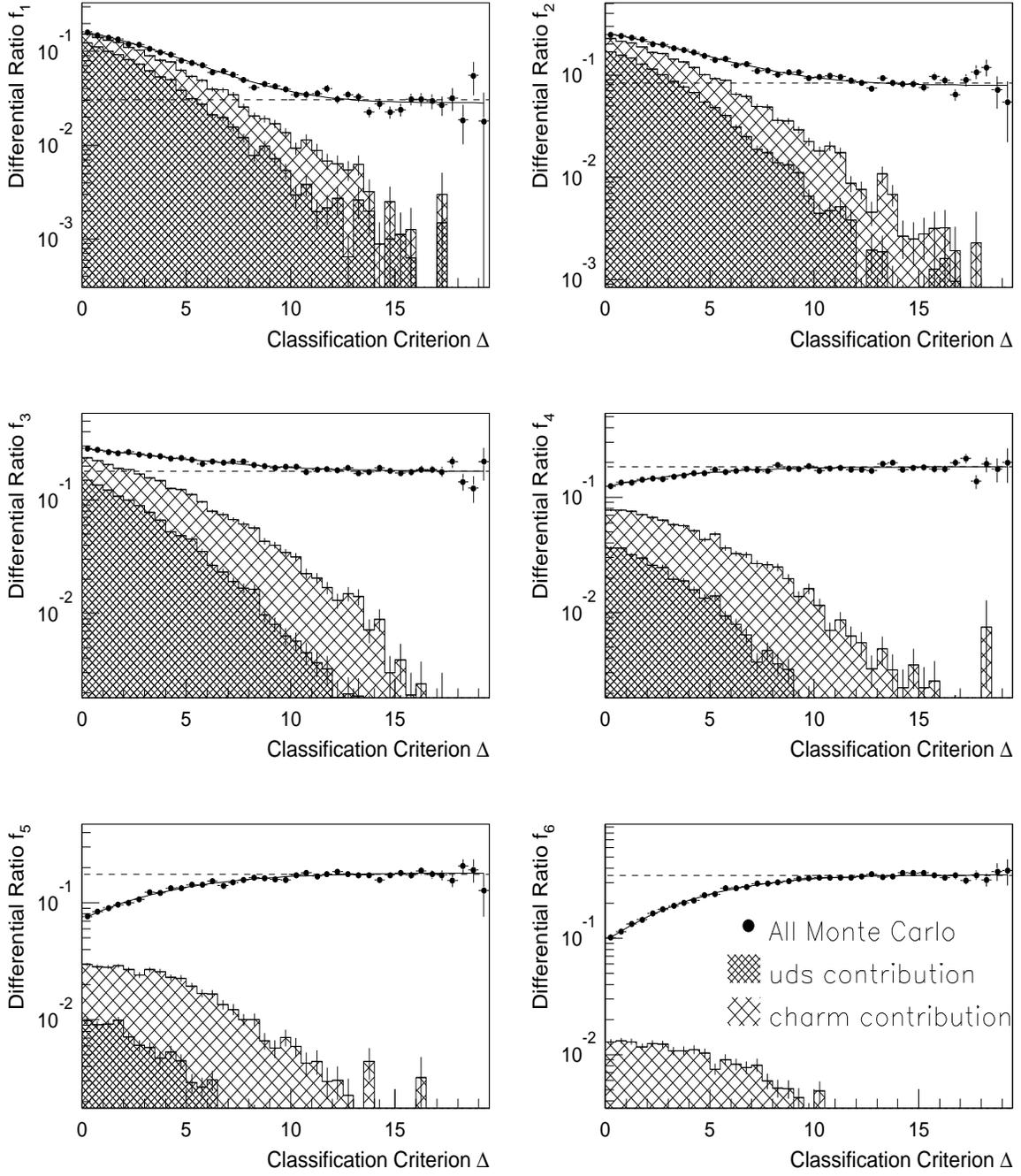
Figure 9: $\hat{f}_I(\Delta)$ *distributions with their asymptotic fits (see text). No irreducible uds and c background is observed in the asymptotic region, specially in $\hat{f}_4$, $\hat{f}_5$ and $\hat{f}_6$ distributions which are the most significant for the R_b extraction. The dashed horizontal line shows the expected value for $C_I^{'b}$.*

Table 1: *Classification matrix elements obtained in the fit and their expected values. The first error is statistical and the second one systematic from an 8% uncertainty in the R_c parameter.*

	expected values		
Category	uds	charm	beauty
$1 = uds - tight$	0.2594	0.1293	0.0304
$2 = uds - loose$	0.3478	0.2409	0.0837
$3 = charm$	0.2928	0.3472	0.1801
$4 = b - loose$	0.0754	0.1622	0.1830
$5 = b - medium$	0.0197	0.0788	0.1755
$6 = b - tight$	0.0049	0.0416	0.3474
	fitted values		
Category	uds	charm	beauty
$1 = uds - tight$	$0.2699 \pm 0.0013 \pm 0.0020$	$0.0948 \pm 0.0053 \pm 0.0072$	0.0283 ± 0.0012
$2 = uds - loose$	$0.3441 \pm 0.0015 \pm 0.0010$	$0.2610 \pm 0.0082 \pm 0.0036$	0.0792 ± 0.0017
$3 = charm$	$0.2856 \pm 0.0016 \pm 0.0010$	$0.3742 \pm 0.0059 \pm 0.0036$	0.1799 ± 0.0024
$4 = b - loose$	$0.0737 \pm 0.0009 \pm 0.0011$	$0.1673 \pm 0.0046 \pm 0.0039$	0.1838 ± 0.0016
$5 = b - medium$	$0.0207 \pm 0.0007 \pm 0.0006$	$0.0692 \pm 0.0046 \pm 0.0021$	0.1794 ± 0.0015
$6 = b - tight$	$0.0060 \pm 0.0010 \pm 0.0004$	$0.0336 \pm 0.0071 \pm 0.0013$	0.3494 ± 0.0028

D_{IJ}^l being the double tag efficiency for the flavour l. These correlation ρ_{JI}^l factors are shown in figure 10 with their statistical errors for the six categories. Most of these factors are small or not significant [6]. The most critical correlation factor for the R_b measurement is ρ_{66}^b. Figure 1.b, introduced for the asymptote method, shows the variation of this coefficient with δ. It has a good behaviour even at large values of δ and for the standard cut its value is $\rho_{66}^b = 0.018 \pm 0.010$.

Correlations appear also at the level of the asymptotic estimates X_I^b and can be formulated as

$$X_I^b = \lim_{\delta \to \infty} f_I(\delta) = \left\{ 1 + \lim_{\delta \to \infty} \xi_I^b(\delta) \right\} C_I^b \tag{37}$$

The $\lim_{\delta \to \infty} \xi_I^b(\delta)$ has been found to be well approximated by ρ_{I6}^b.

We have studied how much R_b changes if one switches on [7] or off the correlation pattern. In the first case one uses the true hemisphere correlation coefficients shown in figure 10 through equations (35) and (37); in the second case these coefficients are neglected in the minimization of the $G_2(C, R)$ objective function. In this application this variation was found to be about $(-0.32 \pm 1.00)\%$ relative to the R_b value. The error is due to the limited Monte Carlo statistics in the determination of the correlation coefficients. This is an indication that the method is almost insensitive to the pattern of correlations. Moreover, asymptotic correlation factors can be changed taking into account the small variations of $\xi_I^b(\delta)$ at large δ. A negligible change, at the level of 5 per mil, was found on the final fitted R_b value.

---

[6] For example, the largest factor is $\rho_{11}^b = 0.68 \pm 0.16$, but it affects only 1/1000 of $b\bar{b}$ events.

[7] The additional constraints $\sum_{IJ} C_I^l C_J^l (1 + \rho_{IJ}^l) = 1$, for each l and $\sum_I C_I^b \left\{ 1 + \lim_{\delta \to \infty} \xi_I^b(\delta) \right\} = 1$ have to be included.
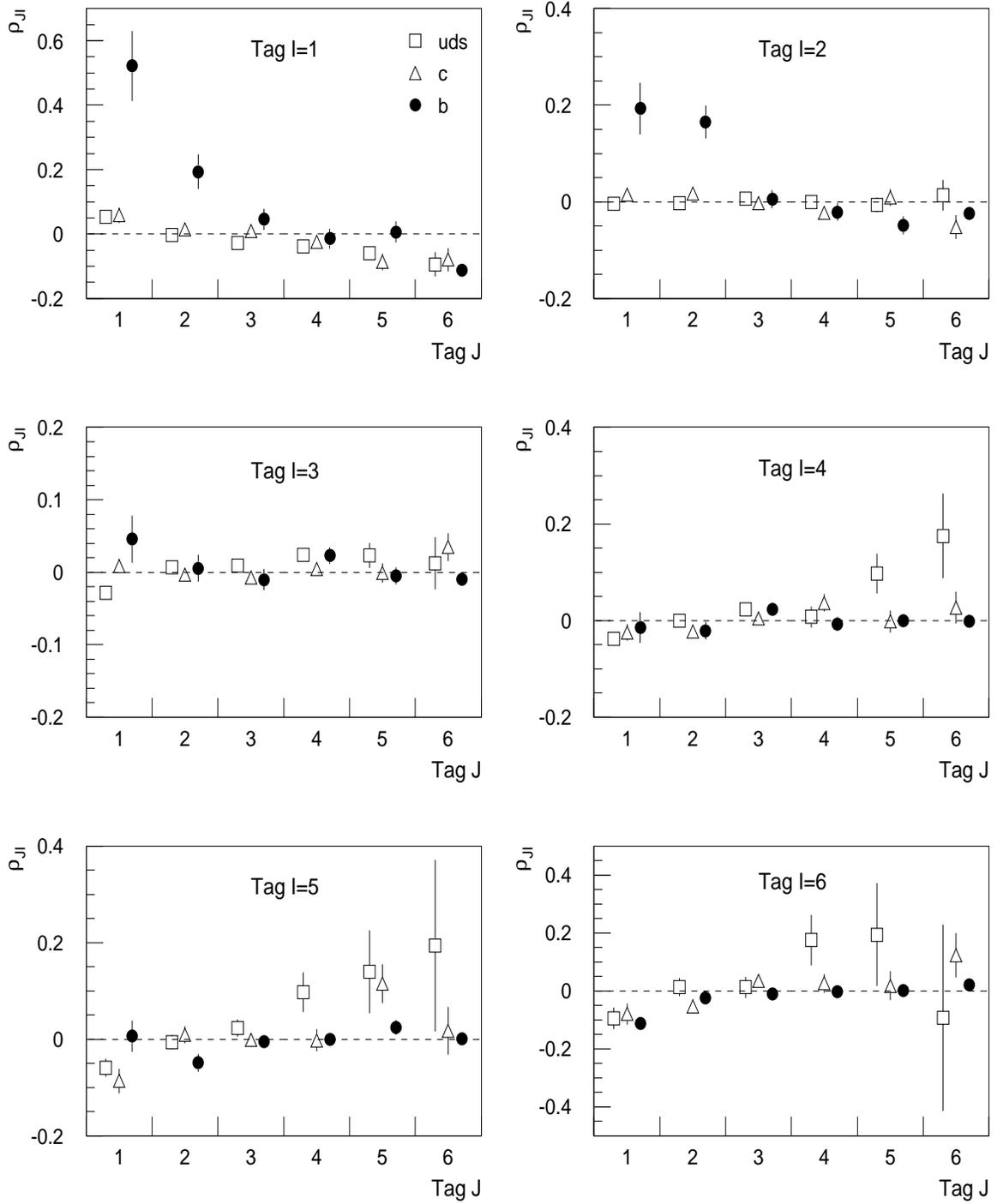
Figure 10: *Double tag hemisphere correlation factors* ρ^l_{JI}.

In the absence of hemisphere correlations, the R_b measurement is mathematically independent of uncertainties on other obervables in and outside the b sector. However, in the presence of small correlations, we have investigated the effect of uncertainties on physical quantities or detector response which may change the correlation pattern and then affect the R_b measurement.

The correlation between hemispheres occurs mainly due to polar angle acceptance, to the fact that the beam spot constraint is common for both hemispheres, and to hard gluon emission that may boost b hadrons into the same hemisphere. Also the correlation pattern may depend on the average b lifetime because it affects the tagging efficiencies.

Hard gluon emission producing a $b\bar{b}$ pair in the same hemisphere is about 2 % of the $b\bar{b}$ events according to the simulation and might be the source of an excess of b events in the *(small I,large J)* and *(large I,small J)* cells. In order to evaluate systematic errors one can perform a fit on simulation, removing the events with two b jets in the same hemisphere and recomputing the b fraction in the reduced sample. The correction to be applied has been found to be around 1% relative.

Generally, methods measuring R_b by hemisphere double tagging require the tag efficiencies for the *uds* and c flavours and take these efficiencies from simulation. They are therefore sensitive to uncertainties outside the b sector. In this method, these efficiencies are measured simultaneously with R_b. In the absence of hemisphere correlations, they should not contribute to systematic errors on R_b. It was verified that these errors due to uncertainties in the *uds* and c sector were of second order. Different lifetimes and relative production rates of D mesons, charm decay multiplicity and fragmentation functions, production rates of long lived particles and secondary interactions were considered.

The sum of all these model uncertainties is at the level of 0.7 % of R_b in this application [2, 3]. This shows that the method is almost insensitive to the uncertainties on the physical parameters.

5.2   Other Systematic Uncertainties

A possible dependence of the value of R_b on the tagging algorithm was investigated. No systematic effects were observed. In fact, when one fits the double tag matrix together with the asymptotic estimation X_I^b, the method calibrates itself since all dependence on the modelling is included in the classification matrix and the R_l parameters are free of this dependence. This behaviour was easily checked by changing directly the b tagging efficiency: when the quality of the tagging improves, the measurement of R_b remains stable while the statistical error decreases because the size of the asymptotical domain is increased. In parallel the method was able to follow accurately the modification of the classification matrix.

We have also investigated the effect of changing: a) the model used to compute the *class likelihoods*, comparing two training samples simulated with different b lifetimes (1.2 and 1.6 ps) and different versions of the simulation program; b) the boundaries δ_b^{low} and δ_b^{up} which define the b categories. A relative error of a 0.30% on the R_b value is found for these effects.

6 Conclusions

Two closely related methods have been proposed to directly extract the flavour contents and the performances of a hemisphere tagging algorithm for samples of events collected in e^+e^- annihilation experiments at LEP/SLC energies. These quantities can be obtained without any explicit reference to information taken from simulation, except eventually residual hemisphere correlations. The methods have been tested on a sample of 540K MC events with a full simulation of the detector, using a sophisticated multivariate analysis technique, optimized for b tagging.

The choice of tagging algorithm is irrelevant, provided that similar performances are reached for the b flavour. Much attention should be paid to reduce as much as possible the correlation between the tags in the two hemispheres. The present work takes benefit of the high precision given by microvertex detectors. It was applied to measure accurately the $Z^0 \rightarrow b\bar{b}$ branching ratio.

The results of the two methods show excellent agreement with the expected values together with a good precision in the b sector. A minimal dependence on modelling has been achieved. Systematic errors due to uncertainties on lifetimes, fragmentation functions, branching ratios, or detector resolution effects have been found small.

The main source of systematics comes from residual hemisphere correlations. Simulation could be used to evaluate these corrections. The Monte Carlo study shows that a global systematic uncertainty at the level of 1.3 % can be achieved on the $\Gamma_{b\bar{b}}/\Gamma_{had}$ branching ratio, where the main contribution ($\sim 1\%$) is due to the limited simulation statistics [2, 3]. Also an accurate determination of the tagging performances provides precise *calibrated* samples of *b-enriched* or *b-depleted* hemispheres.

From the detector point of view, a set of requirements is needed [6]. For instance, good quality of the tracking devices, in particular for the vertex region, is essential. Accuracy of the measurement is directly related to the size of the asymptotical domain. In this respect, the new generation of 3-D vertex detectors, which should improve considerably the quality and the solid angle of the b-tag, offers promising perspectives. Finally, progress needs to be made in a better separation of the c and uds flavour. For further developments it should be interesting to have particle identification fully available, which would allow the introduction of independent discriminating variables.

APPENDIX

A ● Tagging variables

Once the hemisphere vertex has been obtained as explained in section 3.1, a set of twelve variables is computed independently in each hemisphere. Assuming that the vertex detector provides accuracy only in the plane perpendicular to the beam axis, we neglect the z track coordinates. Essential ingredients in these variables are the impact parameter of charged trajectories in the xy plane and related magnitudes. Some nomenclature is first briefly reviewed:

- h_i is the impact parameter of the i-th particle trajectory projected in the xy plane with respect to the hemisphere vertex A_h. The sign of h_i is positive if the vertex A_h is seen on the left when moving along the particle trajectory.

- $\sigma_{h,i}$ is the error on h_i, which adds in quadrature the contribution of the perigee parameter and the error from the vertex position.

- In the xy plane the projection of the impact parameter on the jet axis is $q_i = h_i \sin \epsilon_i$ and normalized to its error

$$\hat{q}_i = h_i \sin \epsilon_i / \sigma_{q,i}$$

where ϵ_i is the angle of the trajectory at perigee with the jet direction. The main error in ϵ_i is due to the estimate of the jet direction. Note that $\hat{q}_i$ is positive for decay products of c and b hadrons travelling in the downstream direction of the jet.

The variables are defined for the N_g *good tracks* which mean charged particles attached to the main interaction or to secondary vertices very close to it. These tracks have to satisfy the following additional conditions:

- The impact parameter should be less than 0.25 cm in the xy plane and less than 1.0 cm in z.

- The impact parameter error must be less than 0.5 cm.

- Tracks within the microvertex acceptance should have at least two associated hits.

In figures 11 and 12 we display the distribution of each variable for *uds*, *c* and *b* flavours, obtained from a simulated sample called the *training sample*. These distributions are used to compute the class likelihoods in equation (29). The figures are plotted with a logarithmic scale, and the contributions of the 3 flavours are on top of each other for readability.

A.1 Boosted sphericity (S_h)

This variable is the only one exclusively computed with quadrimomenta. The jet sphericity of the particles is evaluated with respect to the rest frame of a *B hadron candidate* moving along the jet direction given by the total momentum of the particles in the jet.
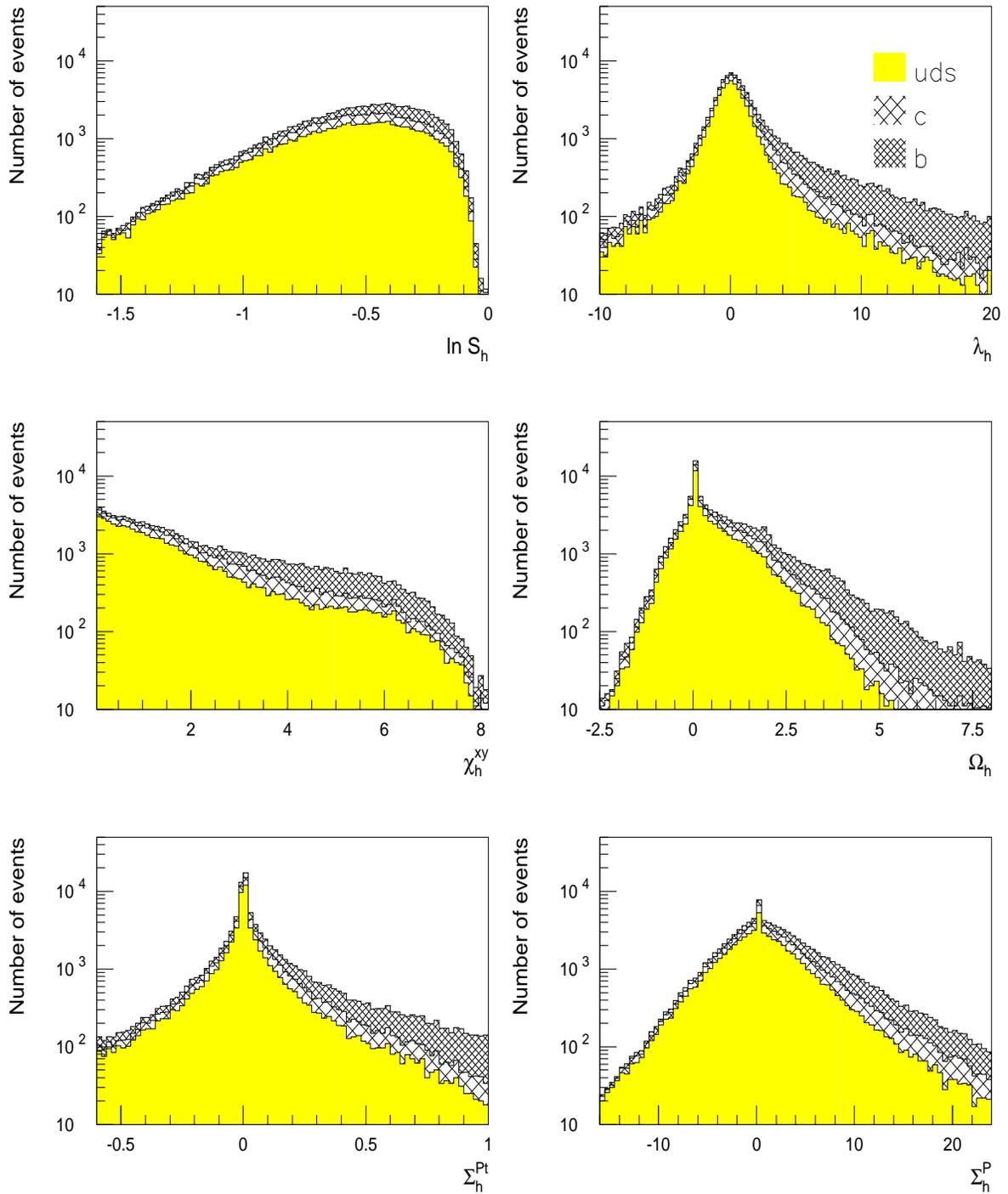
Figure 11: *Distributions for tagging variables (see text). The contribution of each flavour is shown.*
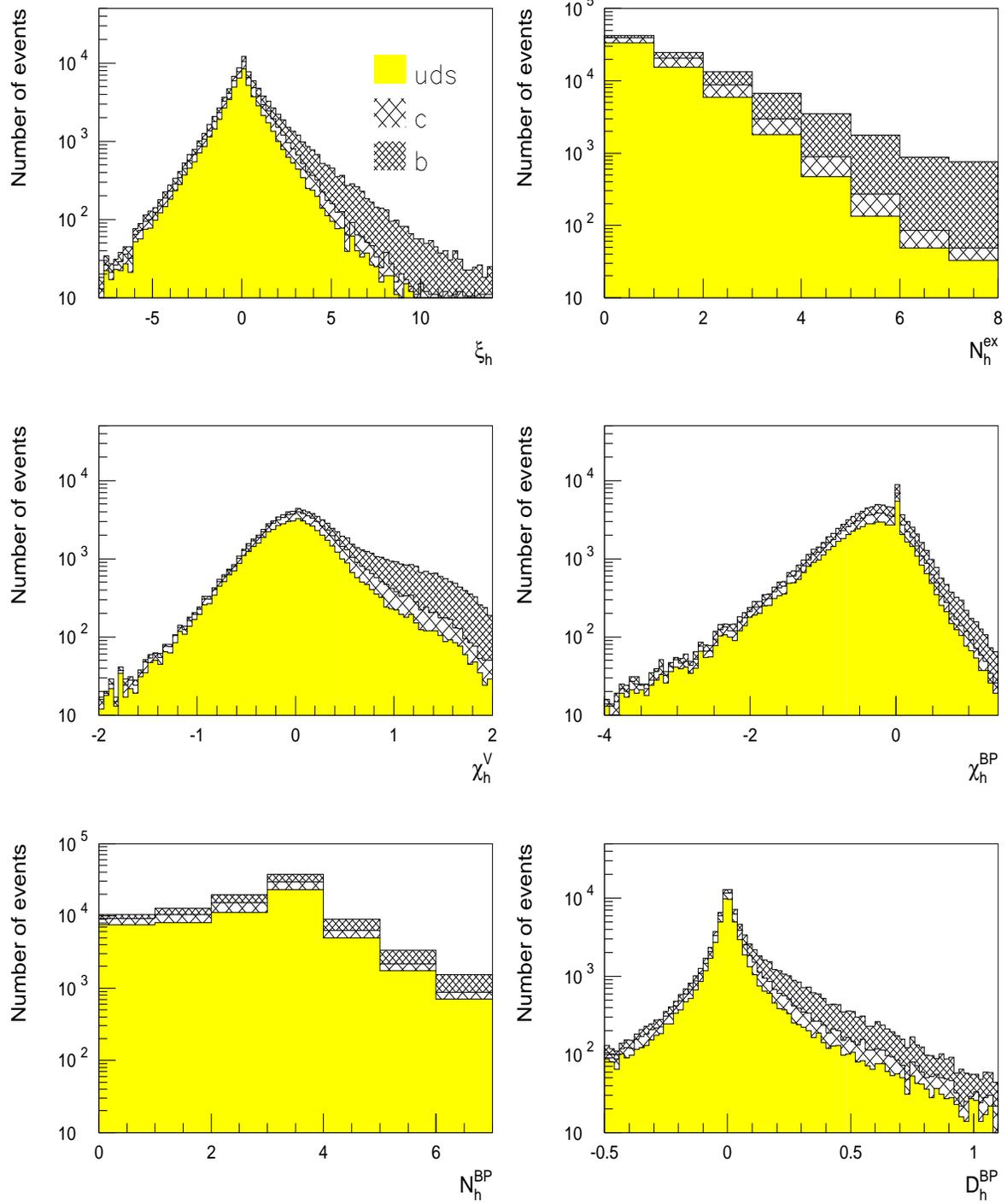
Figure 12: *Distributions for tagging variables (continuation).*

A boost, along the jet direction, with a Lorentz γ parameter is needed to perform the transformation from the laboratory frame to the B rest frame. Monte Carlo studies show that at LEP/SLC energies the optimum value is $\gamma \simeq 4$. The sphericity in this frame is expected to be larger for $b\bar{b}$ events than for the other flavours.

A.2 Normalized decay path (λ_h)

In addition to the main hemisphere vertex, a *secondary* vertex fit is attempted for each hemisphere. The most energetic jet of the hemisphere is associated to the primary quark jet. Only particles making an angle of less than 14° with the jet axis, with more than 1 GeV/c and with an impact parameter respect to A_h of less than 2 mm in xy and 10 mm in z are considered in the secondary vertex fit. The fit provides the position of the secondary vertex V_h and its covariance matrix. If there is only one track remaining in the fit, the secondary vertex is taken as the intersection in the xy projection of this track and the jet axis passing through the main hemisphere vertex A_h. If no track is left to compute the fit, the same procedure is applied to the most energetic remaining jet.

An algebraic distance L_h along the jet direction $\vec{j}_h$ is defined for each hemisphere

$$L_h = \overrightarrow{A_h V_h} \bullet \vec{j}_h$$

and dividing by its error σ_{L_h} the normalized *decay path* variable λ_h is

$$\lambda_h = L_h / \sigma_{L_h}$$

A.3 Sum of impact parameters squared (χ_h^{xy})

By considering the sum of the squared normalized impact parameters with respect to the hemisphere vertex in the xy plane

$$Y_{xy}^2 = \sum_{i=1}^{N_g} \left(\frac{h_i}{\sigma_{h,i}} \right)^2$$

a *pseudo-χ^2* variable can be introduced for each hemisphere. Taking a logarithm scale

$$\chi_h^{xy} = \ln \left(1 + \frac{Y_{xy}^2}{d.o.f.} \right)$$

A.4 Total weight variable (Ω_h)

Among the *good tracks* it is interesting to count the *secondary* tracks coming from c and b decays. Instead of selecting the candidates by a set of cuts, a weight is assigned to each particle giving its probability to be secondary. A weight ω_i is assigned to each track

$$\omega_i = \left\{ 1 - \exp\left(-P_i\right) \right\} \left\{ 1 - \exp\left(-h_i^2/\sigma_{h,i}^2\right) \right\} \tanh\left(\hat{q}_i\right)$$

The weight ω_i is designed to be ≈ 0 for primary tracks (low momentum P_i or $h_i/\sigma_{h,i}$, $\hat{q}_i$ small). The sign introduced by the $\tanh(\hat{q}_i)$ factor allows a cancellation in the summation. On the contrary for secondary tracks ($h_i/\sigma_{h,i}$ large, $\hat{q}_i$ large and positive) ω_i reaches the

value 1. The sum Ω_h of the weights defined above is then equivalent to the *number of secondary particles* in the hemisphere and can be expressed as

$$\Omega_h = \sum_{i=1}^{N_g} \omega_i$$

A.5   Sum of weighted P_T ($\Sigma_h^{P_T}$)

This is another weighted variable equivalent to the *sum of the P_T^2* of secondary particles

$$\Sigma_h^{P_T} = \sum_{i=1}^{N_g} \omega_i P_{T,i}^2$$

A.6   Sum of weighted P (Σ_h^{P})

This variable is defined as the sum of the momenta weighted by the signed factor $\tanh(\hat{q}_i)$

$$\Sigma_h^{P} = \sum_{i=1}^{N_g} \tanh(\hat{q}_i) P_i$$

It intends to represent the sum of the secondary particle momenta. The contribution of primary particles cancel on average.

A.7   Sum of projected impact parameter (ξ_h)

The sum of the projected impact parameters in the xy plane of all *good tracks* is defined as

$$\xi_h = \sum_{i=1}^{N_g} \hat{q}_i$$

The ξ_h distribution is expected to be centered at 0 for the *uds* flavours while for c and b an asymmetry in the positive direction is expected, due to the fact that in general the decay products have positive projected impact parameter.

A.8   Excluded particles (N_h^{ex})

N_h^{ex} is the number of excluded particles during the main hemisphere vertex fit as described in section 3.1.

A.9   Hemisphere vertex χ^2 (χ_h^{V})

The quantity

$$\chi_h^{V} = \ln \frac{\chi_h^2}{d.o.f.}$$

takes into account the χ_h^2 of the xy plane vertex fit for each hemisphere, once the *badly* measured tracks have been removed by applying a χ^2 cut on the trajectory fit.

A.10 Best Partition χ^2 $\left(\chi_h^{BP}\right)$

This is the minimum value of the quantity

$$\chi_h^{BP} = \ln \frac{\chi_{h,1}^2}{d.o.f_1} + \ln \frac{\chi_{h,2}^2}{d.o.f_2}$$

which is evaluated by comparing all possible partitions of the tracks in one hemisphere into two subsets. The $\chi_{h,1}^2$ and $\chi_{h,2}^2$ are the χ^2 values from the vertex fits of the two subsets of tracks. In subset 2, we put up to 6 tracks lying on a cone of 25° around the axis of the most energetic jet in the hemisphere. This selection is sensitive to secondary tracks from b decays. In subset 1, which intends to represent the primary interaction, we put the remaining tracks. The aim of this procedure is to provide a test of the hypothesis that *there are two vertices* in the hemisphere and also an estimation of the secondary vertex (neglecting the cascade $b \rightarrow c \rightarrow s$).

A.11 Best partition tracks $\left(N_h^{BP}\right)$

This is the number of tracks in the subset 2 for the *best partition*.

A.12 Best partition distance $\left(D_h^{BP}\right)$

This is the distance (projected onto the thrust axis) between the fitted vertices from the two subsets in the *best partition*.

B • Glossary of the main mathematical symbols

	GENERAL SYMBOLS
N_F	Number of flavour families considered, here 3: uds, c and b.
l	Flavor index (vary from 1 to N_F).
R_l	Flavour fraction in the sample after acceptance cuts.
Δ	Classification criteria: defined in section 3 as the *winning margin*.
δ	Value of purification cut applied on the classification criteria: $\Delta > \delta$.
	CURVE METHOD SYMBOLS
$S_b(\delta)$	Probability that an hemisphere has a b tag and fulfills the $\Delta > \delta$ cut.
$P_l(\delta)$	Same as $S_b(\delta)$ but for flavour l. $P_b(\delta)$ can be evaluated assymptotically.
$D_{bb}(\delta)$	Probability that an event has a b tag with $\Delta > \delta$ in the two hemispheres.
$\hat{S}_b(\Delta),\hat{D}_{bb}(\Delta)$	Same as $S_b(\delta),D_{bb}(\delta)$, but differential probabilities.
$\hat{P}_b(\Delta)$	Differential probability in the b tag for the b flavour.
$\hat{r}_b(\Delta)$	Differential ratio of single and double b tags tending towards R_b for large Δ.
$\hat{s}_b(\Delta)$	Single b tag ratio requiring an estimate of $\hat{P}_b(\Delta)$,
	tending also asymptotically towards R_b.
$\hat{d}_{bb}(\Delta)$	Same as previously but with double tags.
	MATRIX METHOD SYMBOLS
N_T	Number of categories, here 6.
I,J	Category index ordered by increasing b purity. Categories $1,2$ and $4,5,6$ are
	subdivisions of the uds and b tags. Category 3 corresponds to the charm tag.
C_I^l	Classification probability for flavour l in category I.
D_{IJ}	Fraction of events observed in category I and J for hemispheres 1 and 2,
	input to the fit of $G(\tilde{C},R)$.
D_{IJ}^l	Same as D_{IJ} but for flavour l only.
T_{IJ}	Prediction for D_{IJ}, function of the C and R matrices.
Θ,Ψ,γ	Angles defining the rotation degeneracy.
$\hat{f}_I(\Delta)$	Fraction of hemispheres tagged in category I when the opposite
	hemisphere is classified b with the winning margin Δ.
X_I^b	Asymptotical estimates of tagging probabilities C_I^b deduced
	from the $\hat{f}_I(\Delta)$ distributions.
G	Function to minimize in order to extract C and R (degenerated solutions).
G_1	Extension of G with the X_I^b estimates in a degeneracy-breaking term.
G_2	Extension of G with $\hat{f}_I(\Delta)$ distributions in the degeneracy-breaking term.
ρ_{JI}^l	Double tag hemisphere correlation matrix for a given flavor l.

References

[1] P. Billoir *et al.* (DELPHI Collaboration), *Measurement of $Z^0 \to b\bar{b}$ Branching Ratio by Hemisphere Double Tagging*, LPNHE 94-05 and contributed paper to the ICHEP 93 Conference, Marseille (France), July 1994.

[2] P. Abreu *et al.* (DELPHI Collaboration), *Measurement of the $\Gamma_{b\bar{b}}/\Gamma_{had}$ Branching Ratio of the Z by Double Hemisphere Tagging*, CERN–PPE/94-131 (1994). Submitted to Zeit. f. Physik C and contributed paper to the ICHEP 94 Conference, Glasgow (Scotland), July 1994.

[3] P. Billoir *et al.* (DELPHI Collaboration), *New Measurement of the $\Gamma_{b\bar{b}}/\Gamma_{had}$ Branching Ratio of the Z with Minimal Model Dependence*, IFIC/94-51 and contributed paper to the ICHEP 94 Conference, Ref. gls0229C, Glasgow (Scotland), July 1994.

[4] D. Batignani *et al.* (ALEPH Collaboration), *1991 IEEE Nuclear Science Symposium*, Santa Fe, IEEE transactions on nuclear science, v. NS 39(4-5), Aug. and Oct. 1992, Vol. 1, 438.
N. Bingefors *et al.* (DELPHI Collaboration), Nucl. Inst. Meth. A 328 (1993) 447.
L3 Collaboration, *Proposal for a Silicon Microvertex Detector for L3*, CERN–LEPC 91-5, LEPC p4-Add.1, April 1991.
P.P. Allport *et al.* (OPAL Collaboration), *The OPAL Silicon Strip Microvertex Detector with Two Coordinate Readout*, CERN–PPE/94–16.
C.J.S. Damerell *et al.* (SLD Collaboration), Nucl. Inst. Meth. A 275 (1989) 484.

[5] A. G. Frodesen, O. Skjeggestad, H. Tofte, *Probability and Statistics in Particle Physics*, Universitetsforlaget 1979.

[6] D. Decamp *et al.* (ALEPH Collaboration), Nucl. Inst. Meth. A 294 (1990) 121.
P. Aarnio *et al.* (DELPHI Collaboration), Nucl. Inst. Meth. A 303 (1991) 233.
P. Adeva *et al.* (L3 Collaboration), Nucl. Inst. Meth. A 289 (1990) 35.
P. Ahmet *et al.* (OPAL Collaboration), Nucl. Inst. Meth. A 305 (1991) 275.
J.D. Fox *et al.* (SLD Collaboration), Nucl. Phys. B 23A (1991) 227.

[7] DELSIM Reference Manual, DELPHI 87-98, Geneva, 1989.

[8] T. Sjöstrand, Comp. Phys. Comm. 39 (1986) 347.
T. Sjöstrand and M.Bengtsson, Comp. Phys. Comm. 43 (1987) 367.
T. Sjöstrand: JETSET 7.3 manual, preprint CERN-TH 6488/92 (1992).

[9] P. Billoir *et al.*, *B-tagging by hemisphere: description of variables and results on Monte Carlo*, Internal Report LPNHE 94–05.

[10] Ch. de la Vaissiere and S. Palma-Lopes in the *AIP Heavy Flavour Workshop Proceeding (1989)*, 440.