DELPHI Collaboration

DELPHI 2002-069 CONF 603
22nd July 2002

A Study of the *b*-Quark Fragmentation Function with the DELPHI Detector at LEP I

G. Barker[1], E. Ben-Haim[2,3], M. Feindt[1], U. Kerzel[1], P. Roudeau[2] L. Ramler[1], A. Savoy-Navarro[3]

(1) Institut für Experimentelle Kernphysik, Universität Karlsruhe
(2) U. de Paris-Sud, Lab. de l'Accélérateur Linéaire (L.A.L.), Orsay and IN2P3-CNRS
(3) LPNHE, Universités de Paris 6-7 and IN2P3-CNRS

Abstract

The nature of b-quark jet hadronisation has been investigated using data taken at the Z^0 peak by the DELPHI detector at LEP in the year 1994. Sophisticated neural networks have been trained, on a sample of simulated inclusive b-decays, to reconstruct the kinematic variables x_B and z that are commonly used inside fragmentation models. A regularised unfolding technique has then been used to extract from data the underlying fragmentation functions from these reconstructed variables. Commonly used models of fragmentation are tested by fitting to the unfolded functions and the dependence of the results as a function of the event thrust is investigated. Working in a conjugate moments space, and using the latest expressions for the perturbative QCD contribution, the non-perturbative QCD component has been fitted to give a consistency check of the unfolded $f(z)$ result.

Contributed Paper for ICHEP 2002 (Amsterdam)

1 Introduction and Overview

The development of a $b\bar{b}$ quark pair into stable particles that enter a detector, can be loosely split into three stages: In the first ("perturbative") stage, the b-quarks radiate gluons which in turn may split into $q\bar{q}$ or gg pairs. For gluon energies above about 1 GeV, perturbative QCD can in principle be used to calculate the expected energy distribution of the b-quarks. In practice either exact QCD matrix elements in second order of perturbation theory or leading log parton shower cascades are employed. The second stage concerns the fragmentation of the quarks into colourless hadronic states and is not calculable within perturbation theory. This stage must therefore be phenomenologically modeled as must the third phase, which describes the prompt decay of excited states, and of longer lived weakly decaying states, into stable particles that enter the detector.

This paper describes the measurement from data of a *fragmentation function* $D_b^B(v)$ that can accurately model stage two in terms of some kinematical variable v. This function should thus be interpreted as the probability that a hadron B is produced with a given v from an initial quark b. In the Lund fragmentation model[1], a colour flux *string* links two partons. The creation of quark pairs from the vacuum is then modeled by breaks in the string and the energy sharing at each break is parameterised by the variable z. For the case of an initial b and $\bar{b}$ quark system in the absence of gluon radiation, z is defined as

$$z = \frac{(E+p_{||})_B}{(E+p)_b} \tag{1}$$

where $p_{||}$ represents the hadron momentum in the direction of the b-quark and $(E+p)_b$ is the sum of the energy and momentum of the b-quark just before fragmentation.

From an experimental viewpoint, z has some drawbacks: it is a quantity that is (a) not directly accessible in a detector and (b) only defined within the context of a particular perturbative and non-perturbative hadronisation scheme. Because of these limitations, another popular choice of variable with which to investigate fragmentation functions is

$$x_B = \frac{2E_B}{\sqrt{s}} \tag{2}$$

where E_B is the reconstructed b-hadron energy and $\sqrt{s}$ is the c.m. energy which are both quantities that can in principle be directly reconstructed by a detector. It is necessary however when discussing x_B to be clear about exactly which b-hadron is being considered. The *primary* b-hadron is the state created directly after the hadronisation phase, whereas the *weak* b-hadron is the state that finally decays somewhere in our detector volume in a flavour-changing process. Primary b-hadrons are either mesons (about 90%) or baryons (about 10%). In the case of mesons, 30% of the time the primary b-hadron is expected to be an orbitally excited B^{**} meson, 52.5% of the time a B^* meson, and in only 17.5% of cases, directly a weakly decaying B^+, B^0 or B_s^0 meson. B^{**} and B^* mesons finally decay via kaon, pion or photon emission into weakly decaying ground state mesons, which then carry less energy than their parents. Fortunately, due to the large mass of the b-quark, the mean energy difference between primary and weakly decaying B hadrons is small but not insignificant.

The current analysis investigates fragmentation functions based on the reconstruction of the primary (x_B^{prim}) and weakly decaying B-hadron energy (x_B^{weak}) as well as, for the

first time, reconstructing the Lund string fragmentation variable z by an inclusive technique employing the use of neural networks. The main challenge to the analysis is the difficult problem of how to unfold from the data the underlying $f(x_B)$ and $f(z)$ distributions. The physics motivation is clearly the insight into non-perturbative and perturbative strong interaction effects that knowledge of fragmentation functions gives and valuable constraints on the validity of various hadronisation models.

After a presentation of details of the DELPHI detector and event selection, Section 4 describes the reconstruction of the kinematic variables x_B and z. Section 5 then goes on to explain how a regularised unfolding technique is applied to the reconstructed quantities, to extract from the data the underlying $f(x_B)$ and $f(z)$ distributions.

The results of the unfolding are presented in Section 6 which includes a study of the dependence as a function of the event thrust i.e. hard gluon radiation. Section 7 covers the determination of systematic uncertainties and Section 8 presents fits of the unfolded distributions to a selection of hadronisation model functions.

In section 9, recent NLL perturbative QCD computations [23] are used in conjunction with the unfolding result for the weakly decaying b-hadron in order to extract the non-perturbative component.

This analysis is based strongly on the DELPHI inclusive b-physics package, BSAURUS[2]. Aspects of BSAURUS directly related to the analysis are presented in a summarised form but the reference should be consulted for full details of the package.

2 The DELPHI detector

A complete overview of the DELPHI detector[3] and its performance[4] have been described in detail elsewhere. What follows is a short description of the elements most relevant for this analysis.

Charged particle tracking was performed by the Vertex Detector (VD), the Inner Detector, the Time Projection Chamber (TPC) and the Outer Detector. A highly uniform magnetic field of 1.23 T parallel to the e^+e^- beam direction, was provided by the superconducting solenoid throughout the tracking volume. The momenta of charged particle tracks were reconstructed with a precision of $\sigma_p/p < 2.0 \times 10^{-3} \cdot p$ (p in GeV/c) in the polar angle region $25^\circ < \theta < 155^\circ$. The VD consisted of three layers of silicon micro-strip devices with an intrinsic resolution of about 8 μm in the $R-\phi$ plane transverse to the beam line. In addition, the inner and outer-most layers were instrumented with double-sided devices providing coordinates of similar precision in the $R-z$ plane along the direction of the beams. For charged tracks with hits in all three $R\phi$ VD layers the impact parameter resolution was $\sigma^2_{R\phi} = ([61/(P\sin^{3/2}\theta)]^2 + 20^2)$ μm^2 and for tracks with hits in both Rz layers and with polar angle $\theta \approx 90^\circ$, $\sigma^2_{Rz} = ([67/(P\sin^{5/2}\theta)]^2 + 33^2)$ μm^2. Calorimeters detected photons and neutral hadrons by the total absorption of their energy. The High-density Projection Chamber (HPC) provided electromagnetic calorimetry coverage in the polar angle region $46^\circ < \theta < 134^\circ$ giving a relative precision on the measured energy E of $\sigma_E/E = 0.32/\sqrt{E} \oplus 0.043$ (E in GeV). In addition, each HPC module worked essentially as a small TPC charting the spatial development of showers and so providing an angular resolution exceeding that of the detector granularity alone. For high energy photons the angular precisions were ± 1.7 mrad in the azimuthal angle ϕ and ± 1.0 mrad in the polar angle θ.

The Hadron Calorimeter was installed in the return yoke of the DELPHI solenoid and provided a relative precision on the measured energy of $\sigma_E/E = 1.12/\sqrt{E} \oplus 0.21$ (E in GeV).

Powerful particle identification was made possible by the combination of dE/dx information from the TPC (and to a lesser extent from the VD) with information from the Ring Imaging CHerenkov counters (RICH) in both the forward and barrel regions. The RICH devices utilised both liquid and gas radiators in order to optimise coverage across a wide momentum range: liquid was used for the momentum range from 0.7 GeV/c to 8 GeV/c and the gas radiator for the range 2.5 GeV/c to 25 GeV/c. To combine the information optimally, algorithms[5] based on neural network techniques were used which resulted in e.g. an efficiency for the correct identification of $K^{\pm}$ of 90%(70%) with a contamination of 15%(30%) for $p < 0.7$ GeV/c ($p > 0.7$ GeV/c).

3 Event Selection

3.1 Selecting Multihadronic Events

The data consisted of events taken at centre-of-mass energies at the Z^0 pole in 1994. Multihadronic Z^0 decays were selected by the following requirements:

- at least 5 reconstructed charged particles,
- the summed energy in charged particles with momentum greater than 0.2 GeV/c had to be larger than 12% of the centre-of-mass energy, with at least 3% of it in each of the forward and backward hemispheres defined with respect to the beam axis.

These requirements resulted in the selection of about 1.36 million events in the data with an efficiency estimated to be 92.6% and all backgrounds below the 0.1% level. The Monte Carlo sample of $Z \to q\bar{q}$ events, details of which are listed in Table 1, was approximately 3 times the statistics of the data. The generated events were passed through a full detector simulation[4] and the same multihadronic selection criteria as the data.

Monte Carlo Generator	JETSET 7.3[6]
Perturbative ansatz	Parton shower ($\Lambda_{QCD} = 0.346$ GeV,$Q_0 = 2.25$ GeV)
Non-perturbative ansatz	String fragmentation
Fragmentation function	Peterson et. al.[7] ($\epsilon_b = 0.002326$)

Table 1: Details of the Monte Carlo generator used together with some of the more relevant parameter values that have been tuned to the DELPHI data.

3.2 Event Hemisphere Selection

Event hemispheres used for the analysis were accepted if the following criteria were fulfilled.

- $|\cos\Theta_{thrust}| < 0.7$.

- The event was tagged as a $Z \to b\bar{b}$ candidate event by the standard DELPHI b-tagging package[9].
- The secondary vertex fit converged successfully according to limits set in BSAURUS on the fit χ^2 value and the number of minimisation iterations needed.
- $0.5 < (x_{hem} = E_{hem}/E_{beam}) < 1.1$

After this selection, 227940 hemispheres remained in the data with a purity (as calculated from the Monte Carlo), in $b\bar{b}$ events of 96%.

4 The Reconstruction of E_B^{prim}, E_B^{weak} and z

The three kinematical variables used to study fragmentation functions, the energy of the weakly decaying B hadron, the energy of the primary B and the LUND string model variable z, were reconstructed using separate artificial neural networks. The ansatz used applied a Bayesian interpretation to the network output in order to return a conditional probability density function for the variable on a *hemisphere-by-hemisphere* basis. The median (and associated error) of the extracted p.d.f. was then defined to be the reconstructed fragmentation variable and its error.

The networks were trained using a list of 22 input variables which included different estimators of the energy available in the hemisphere together with some measures of the expected quality of such estimators e.g. hemisphere track multiplicity and hemisphere reconstructed energy. The inputs were the same for each of the three networks trained, the only difference being the target value i.e. E_B^{prim}, E_B^{weak} or z respectively. The degree of correlation of the inputs to the network target value naturally varies from case to case. A preprocessing stage to the network algorithm automatically suppresses the influence of the inputs with low correlation to retain optimal performance. For completeness, the full list of variables input to the networks is given below (note that some variables have been summarised into a single item in this list):

- Rapidity[1] of the track with the highest value in the event hemisphere.
- Rapidity of the track (see Figure 1) with the second highest value in the event hemisphere.
- The energy component of the TrackNet[2] weighted sum of 4-momenta for the case of 2-jet events. For the case of $>$ 2-jet events, the energy component of the 4-vector as reconstructed by the *rapidity algorithm* is taken. The rapidity algorithm sums over all particles in a hemisphere that pass a cut of $y > 1.6$. This procedure defines an estimate of the weakly decaying B-hadron 4-vector, P_{raw}.
- Energy of the weakly decaying B-hadron as estimated by the rapidity algorithm.
- Mass of the weakly decaying B-hadron as estimated by the rapidity algorithm.

[1] Defined as $y = \frac{1}{2} \cdot \log\left((E + P_\parallel)/(E - P_\parallel)\right)$ where $P_\parallel$ is the momentum component of the track in the direction of the b-quark. The direction is estimated as the axis of the jet associated with the b-hadron.

[2] The TrackNet is a neural network trained to distinguish between charged particles from the B hadron decay chain and those originating from the event primary vertex. See Figure 1.

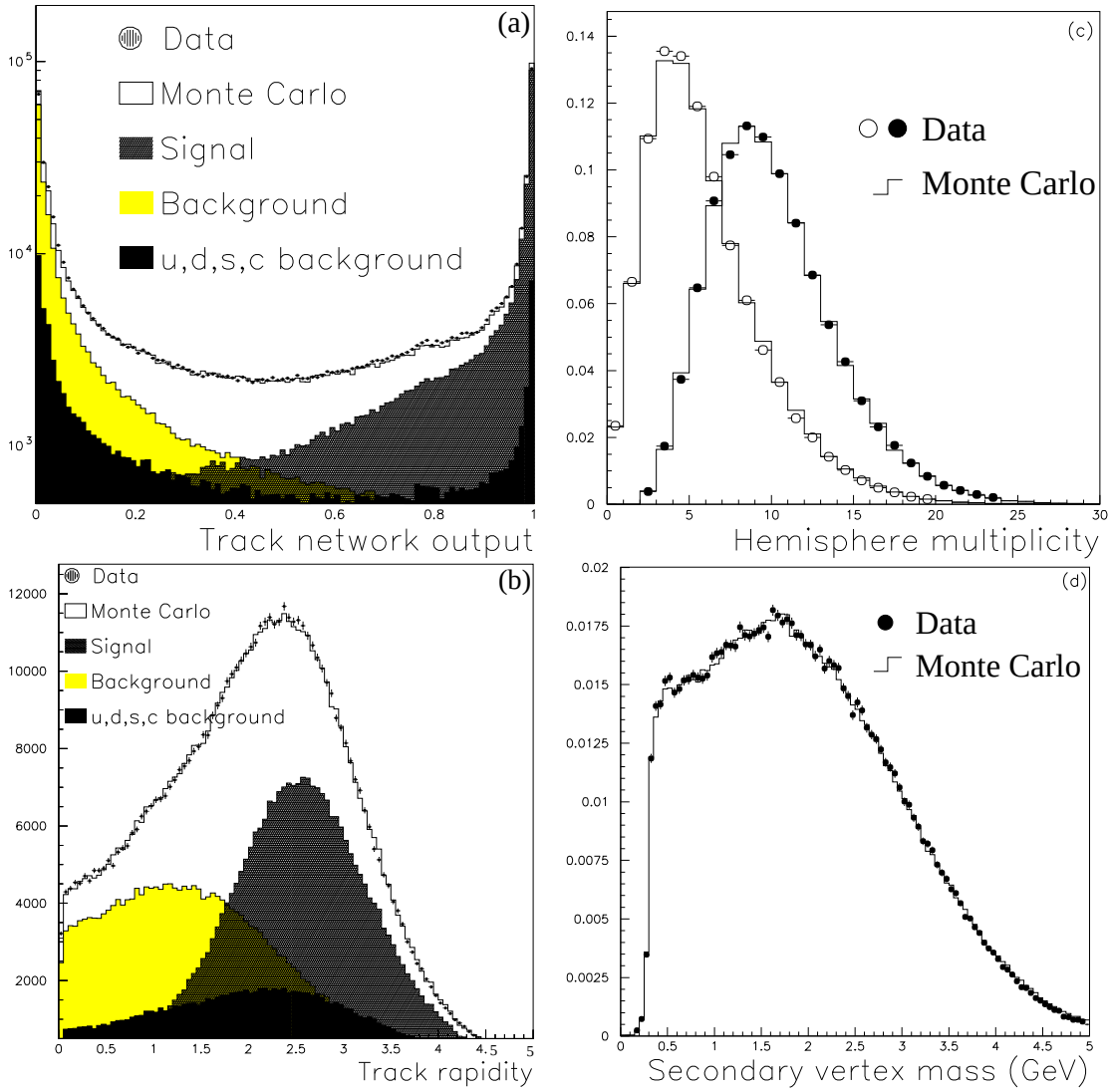

Figure 1: A selection of input variables to the neural networks showing Monte Carlo compared to data. (a) TrackNet: where 'signal' refers to tracks that originate from the B-decay chain and 'background' are tracks in $b\bar{b}$ events not from the B-decay chain; (b) the rapidity of charge tracks with the same definition of signal and background as for the TrackNet; (c) charged particle multiplicity distributions where the closed points represent the total particle (neutral and charged) multiplicity in the hemisphere and the open points show the difference between the number of hemisphere tracks passing the selection cuts and the number that, in addition, pass a TrackNet cut of 0.5; (d) the invariant mass of all tracks associated with the secondary vertex fit in $b\bar{b}$ tagged events.

- Energy of the jet with the highest $b-$tag value.
- Total energy in the hemisphere scaled by the beam energy.
- The total hemisphere energy normalised by an estimate of the c.m. energy $E_{c.m.}$, given by considering the Z^0 to decay into the two-body final state of a B-jet with mass M_{B-jet} recoiling against all other particles in the event with mass M_{recoil} i.e. in the rest frame of the Z^0,

$$E_{c.m.} = \frac{M_{Z^0}^2 - M_{recoil}^2 + M_{B-jet}^2}{2 \cdot M_{Z^0}} \tag{3}$$

- An estimate of $E_{c.m.}$ as given by Eqn. 3 where M_{recoil} is based on charged tracks only. In the hemisphere containing the B-candidate, only tracks likely to have originated from the fragmentation process are selected (by use of the TrackNet) whereas all tracks are used in the opposite hemisphere.
- A further estimate of $E_{c.m.}$ as given by Eqn. 3 where neutrals from the opposite hemisphere are also included in the formulation of M_{recoil}.
- An estimate of the missing p_T between the B-candidate direction and the thrust axis calculated using only fragmentation tracks in the same hemisphere as the B-candidate (via use of the TrackNet) and all tracks in the opposite hemisphere. The calculation is repeated for two different definitions of the B-direction: one being the vector pointing from reconstructed primary vertex to reconstructed secondary vertex and the other given by the vector $\underline{P}_{raw}$.
- The mass of the reconstructed secondary vertex. See Figure 1.
- The polar angle of the B-candidate momentum vector.
- The thrust value of the event.
- The difference between the number of tracks in the hemisphere passing the selection cuts and the number of such tracks that, in addition, pass a TrackNet cut of 0.5. See Figure 1.
- The total number of all charged and neutral particles. See Figure 1.
- The number of particles passing a TrackNet cut of 0.5.
- The probability that the best electron or muon candidate in the hemisphere, with the correct charge correlation, originates from the B-candidate.
- The gap in rapidity between the track of highest rapidity with TrackNet value less than 0.5 and the track of lowest rapidity and TrackNet value greater than 0.5.
- The (binomial) error of the vertex charge measurement, Q_v, defined as,

$$\sigma_{Q_v} = \sum_{i}^{tracks} \sqrt{P_B(i)(1 - P_B(i))} \tag{4}$$

  where $P_B(i)$ is the TrackNet value for track i.

- A hemisphere 'quality flag' constructed by counting the number of tracks in the hemisphere likely to be badly reconstructed i.e. the number failing the selection cuts of Section 3.

The precision of the estimators is shown in Figure 2 after all analysis selection cuts have been applied and based on a statistically independent Monte Carlo sample to that used for training.

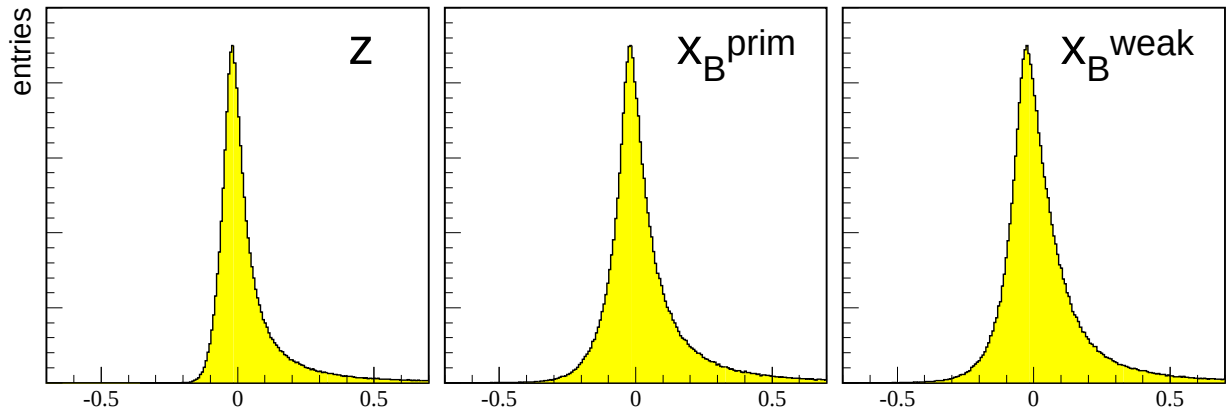

Figure 2: Plots showing the resolution i.e. $(v_{rec} - v)/v$ for $z(FWHM = 7.7\%), X_B^{prim}(FWHM = 11.2\%)$ and $v = X_B^{weak}(FWHM = 14.0\%)$.

5 The Unfolding Method

The experimental challenge of the analysis is to determine from the measured distribution $g(v_{rec})$ in data, where $v = x_B^{prim}, x_B^{weak}$ or z, the underlying physics probability density function $f(v)$. In general $g(v_{rec})$ will differ from $f(v)$ due to:

- finite detector resolution
- limited measurement acceptance
- variable transformation i.e. any biases or distortions that may be present in the measured quantity.

Mathematically, the distributions are related by:

$$g(v_{rec}) = \int R(v_{rec}; v) f(v) dv + b(v_{rec}), \tag{5}$$

where $b(v_{rec})$ is the background contribution (in our case mainly non-$b\bar{b}$ events) to the measured distribution and is taken from Monte Carlo simulation. The *response function* $R(v_{rec}; v)$ provides the mapping of true v to v_{rec} and thus contains all the effects of resolution, acceptance and transformation mentioned above.

In order to *unfold* the physics distribution $f(v)$ from Equation 5 we used the RUN (**R**egularized **UN**folding) program[10]. The algorithm defines a function $W(v)$ used to

provide a weight to the Monte Carlo distribution $f(v_{rec})_{MC}$ such that it reproduces the data distribution $g(v_{rec})$ as well as possible i.e. $W(v)$ is determined by a fit to the data. The result of the unfolding, up to a normalisation factor, is then given by

$$f(v) \;=\; W(v) \cdot f_{MC}(v) \tag{6}$$

and by integration over bins in v, unfolded binned points are determined together with a complete covariance matrix.

The binning of the unfolded result is an important aspect of the RUN procedure, which is determined ultimately by the experimental resolution. To see how this comes about, we must consider the internal representation of the weight function namely,

$$W(v) = \sum_{j=0}^{n_j} a_j \cdot p_j(v) \tag{7}$$

where the $p_j(v)$ form a normalised, orthogonal set of polynomials[3] where $p_0(v)$ is usually close to a constant, $p_1(v)$ is linear in v , $p_2(v)$ is quadratic in v , and so on. In practice it is usually found that only the first few of the coefficients, a_j, are significantly different from zero and so contributions from higher coefficients do not add any information to the result. In fact these higher order contributions are damaging since they add oscillating terms to the result and so should be removed. The order at which to terminate the expansion in Equation 7, $j_{cut} < n_j$, is mainly determined by the experimental resolution. This cut-off then also fixes the rank of the subsequent covariance matrix, and hence the number of unfolded bins, to be j_{cut}. It is important to realise that an arbitrarily large number of unfolded bins can always be taken but at the expense of bin-to-bin correlations becoming very high and the covariance matrix becoming singular, since the rank will remain j_{cut}. The expansion cut-off inside RUN is actually implemented not as a sharp cut-off point but as a smooth transition to zero via a regularisation procedure. Further details of the unfolding method can be found in [11].

6 Unfolding Results

6.1 Crosschecks

As a check of the method principle, a sample of the Monte Carlo events were used as real data with the remainder used as simulation and the unfolding procedure was run as normal. The result for the case of unfolding x_B^{prim} and x_B^{weak} is shown in Figure 3 and indicates an excellent reproduction of the overall shape and normalisation of the underlying true distribution (labeled as 'MC truth'). The correlation between the reconstructed and the true quantity for $x_B^{prim}(_B^{weak})$ was estimated from the Monte Carlo to be 82%(82%). Figure 3 illustrates that the procedure works equally well for the case of unfolding z where the correlation between the measured and the true distributions is 38%. This test confirms that the method is rather robust in the calculation of the response function when the measured distribution has only a rather weak correlation to the true underlying physics distribution.

[3]Parameterised by B-splines.

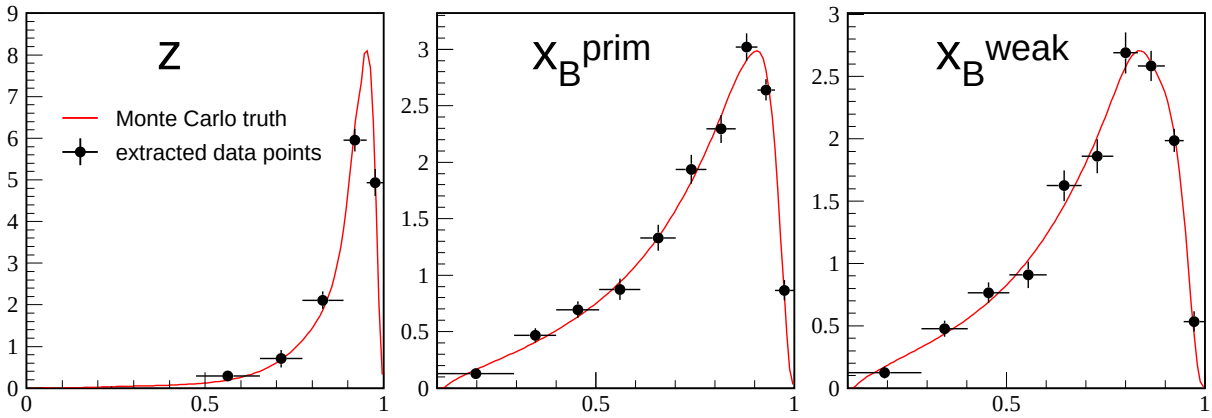

Figure 3: The result of the unfolding procedure run on Monte Carlo data for the case of unfolding z, x_B^{prim} and x_B^{weak}.

By construction, the unfolding procedure should be largely independent of the *prior* fragmentation function, $f(v)_{MC}$, which was used to generate the Monte Carlo sample. To test the validity of this claim for the case of unfolding z, the data was represented by the default Monte Carlo sample generated with the Peterson function. The statistically independent Monte Carlo sample used to estimate the transfer function $R(y, z)$ was however re-weighted such that the underlying true distribution was given by an arbitrary function (taken to be $f(z) = z(1-z)\exp(3z)$). The result of the unfolding is shown in Figure 4 and illustrates that, even for this extreme example, the correct underlying functional form is reproduced.

6.2 Results From Data

The results of the unfolding applied to the real data set are displayed in Figure 5 for the case of z, x_B^{prim} and x_B^{weak}. It should be noted that for the case of z it is found to be important that the weighting function, $W(v)$ of Equation 6, is constrained to be positive. This ensures that the resulting unfolded function does not move into the unphysical negative region so allowing the function to be interpreted as a probability density distribution. The plots show the unfolded distributions and the corresponding binned data points, as defined above, together with an overlay of the truth input distributions for comparison. In addition, Figure 6 shows distributions of z, x_B^{prim} and x_B^{weak} in the data compared to Monte Carlo appropiately weighted for the results of the unfolding.

In order to quantify the shapes of the unfolded distributions, Table 2 presents the mean (defined as, $\langle z \rangle = \int_0^1 z f(z) dz$) and variance (defined as, $V(z) = \int_0^1 (z - \langle z \rangle)^2 f(z) dz$) for each of the three cases. The full bin-to-bin results including covariance matrices are listed in Appendix A.

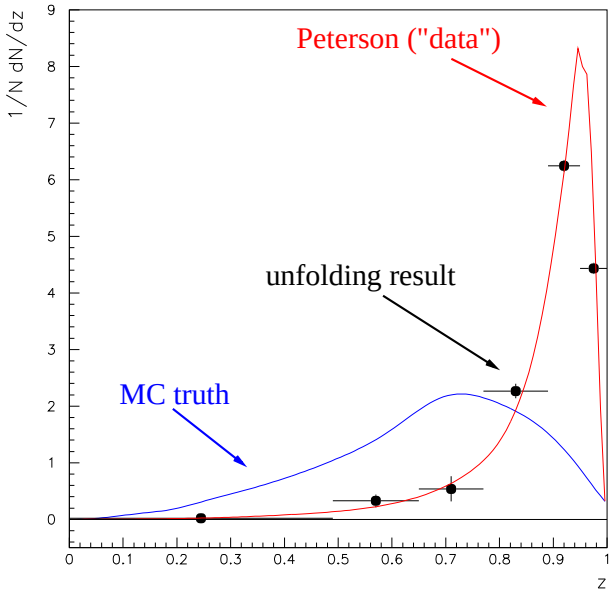

Figure 4: The result of the unfolding where a sample of the default Monte Carlo has been taken as 'data' and an independent Monte Carlo sample has been re-weighted for an arbitrary functional form of the prior distribution. The Peterson function which was used to generate the 'data' sample is accurately unfolded. This example is based on z but the same conclusion holds for unfolding x_B.

	Mean	Variance
$f(x_B^{prim})$	$0.7346 \pm 0.0008(stat.)$	$0.0318 \pm 0.0005(stat.)$
$f(x_B^{weak})$	$0.7153 \pm 0.0007(stat.)$	$0.0300 \pm 0.0004(stat.)$
$f(z)$	$0.8872 \pm 0.0012(stat.)$	$0.0027 \pm 0.0001(stat.)$

Table 2: The means and variances of the unfolded distributions in x_B^{prim}, x_B^{weak} and z.

Note that the mean values for $f(x_B^{prim})$ and $f(x_B^{weak})$ have been corrected to account for the (small) effect of Initial State Radiation (ISR) on our definition of x. In the analysis x is formed by scaling E_B^{prim} and E_B^{weak} by the nominal beam energy of 45.6 GeV and is only strictly correct in the case of no ISR. In $\sim 10\%$ of cases ISR would reduce the energy available for the fragmenting b-quark system from the nominal 45.6 GeV. The size of this effect on the analysis was evaluated from the Monte Carlo.

The results show that there is a disagreement in shape between the distributions unfolded from data and the probability functions used in the DELPHI Monte Carlo generator. The discrepancy becomes most pronounced for the case of z.

Figure 7 presents a scan of the mean of the unfolded distributions in x_B^{weak} and z as a function of the event thrust. This dependence is of interest phenomenologically, particularly for the case of z, since it charts how the underlying energy sharing between quarks and hadrons varies as a function of the (hard) gluon content of the event. As expected, $\langle z \rangle$ is consistent with being independent of thrust, whereas $\langle x_B \rangle$ clearly shows a dependence with hard gluon radiation.

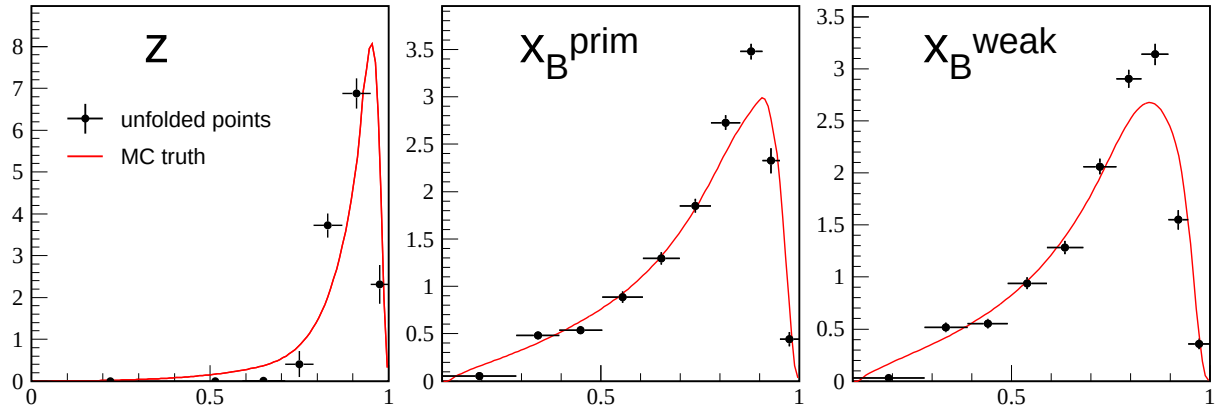

Figure 5: The result of unfolding z, x_B^{prim} and x_B^{weak} from real data. The points represent the unfolding results and the overlayed histograms show the input truth Monte Carlo distributions. The errors are statistical only.

7 Systematic uncertainties

Systematic uncertainties have been evaluated from a wide variety of sources and are presented in Table 3 as errors on the mean value of the unfolded fragmentation functions. Errors, statistical and systematic, for each unfolded bin of the distributions are given in Appendix A together with the associated covariance matrices. Elements of the full covariance matrix have been constructed assuming the systematic errors are 100% correlated bin-to-bin in the following way:

$$cov_{ij} = S_i \cdot S_j + \rho_{ij} \cdot \sigma_i \cdot \sigma_j \tag{8}$$

where cov_{ij} is the ij^{th} element of the covariance matrix, $S_i(\sigma_i)$ are the systematic(statistical) errors associated with bin i and ρ_{ij} is the statistical correlation between bin i and j as given by the unfolding procedure.

7.1 Technical Systematics

Some crosschecks of the unfolding method have been presented earlier in Section 6.1. In addition, an investigation was made of the sensitivity to the following technical aspects of the RUN unfolding procedure:

- The number of unfolded data points (or number of degrees of freedom) was varied and an error assigned based on the spread in the results seen.
- The number of knots in the B-spline parameterisation of Equation 7 was varied and an error assigned based on the spread in the results.
- The binning of the reconstructed variable from Monte Carlo should be well matched to the resolution achieved in order to use the information optimally. A wide range of different binnings, around the default choice, was investigated and the results were

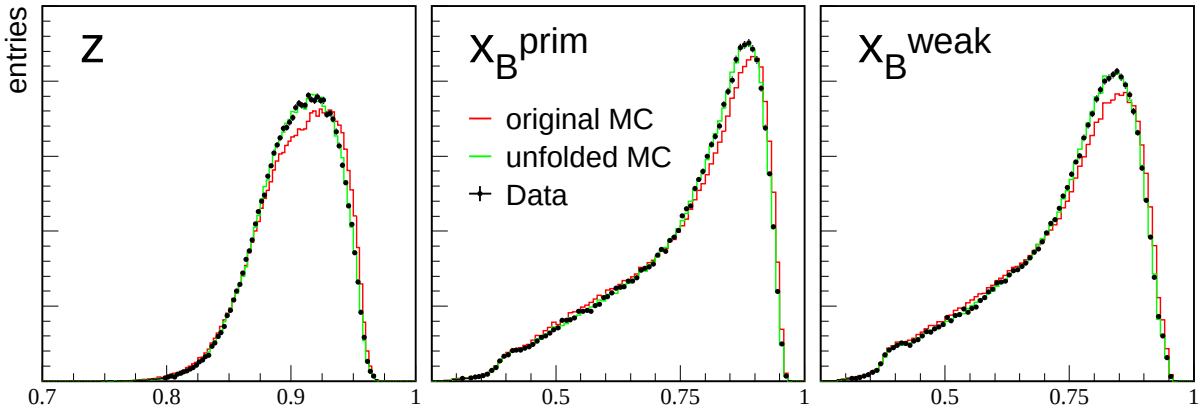

Figure 6: Distributions of z, x_B^{prim} and x_B^{weak} in the data compared to both the default Monte Carlo and the Monte Carlo weighted for the results of the fragmentation function unfolding.

consistent within the errors. Also, no improvement on the statistical precision was found. No additional systematic error was assigned.

7.2 Cuts and Background Dependence

The hemisphere selection described in Section 3.2, includes cuts for $b\bar{b}$ event enhancement and on the reconstructed hemisphere energy x_{hem}, both of which could potentially have an effect on the analysis if not accurately modeled in the Monte Carlo.

The DELPHI b-tagging is based on impact parameter measurements which degrade at low momenta due to the increased effects of multiple scattering. This effect correlates the b-tag information to the B energy and in Figure 8 we present the variation in the unfolding result for $\langle z \rangle, \langle x_B^{prim} \rangle$ and $\langle x_B^{weak} \rangle$ scanned over a wide range of b-tagging cuts i.e. different $b\bar{b}$-purities. The results were found to be stable around the working point of $b\bar{b}$-purity $\approx 96\%$ and no explicit systematic was assigned due to this source.

The effect of scanning around the working point cut value of $x_{hem} = 0.5$ was also investigated and the results presented in Figure 9. The stability around the working point was deemed to be good enough that no explicit systematic was assigned due to this source.

Uncertainties in the size and composition of the background, i.e. $b(v_{rec})$ in Equation 5, were also evaluated. Approximately 75% of the background was from non-$b\bar{b}$ events and is considered in the next Section 7.4. The remainder was composed of cases where both b-quarks were found in the same hemisphere (due to e.g. severe three-jet events) or where a gluon had split into two b-quarks leaving a topology with four b-quarks in the initial state. In these cases, which occur in about 2% of all hemispheres, the connection between the generated b-hadron energy and the reconstructed quantity becomes confused and hence were assigned to the background. We assume that the jet rate is well modeled in the Monte Carlo but vary the gluon splitting rate to $b\bar{b}$ (by default, 0.5%) by $\pm 100\%$ and record the change seen in the unfolding result as a systematic error.

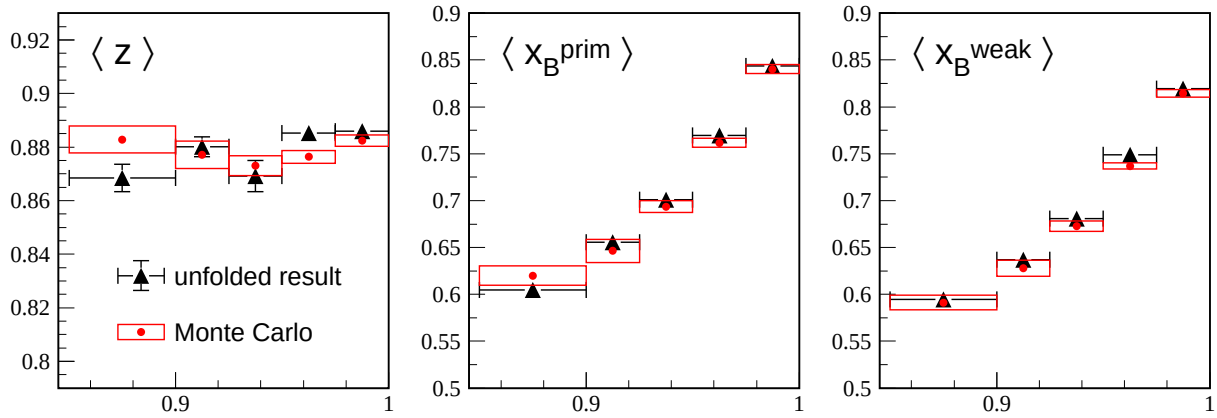

Figure 7: The variation in the mean of the unfolded distributions for $\langle z \rangle$, $\langle x_B^{prim} \rangle$ and $\langle x_B^{weak} \rangle$ in bins of the event thrust.

7.3 Reconstructed Energy

The relationship between the reconstructed variable distribution in the Monte Carlo, $g(v_{rec})_{MC}$, and the underlying physics p.d.f., $f(v)_{MC}$, is

$$g(v_{rec})_{MC} = \int R(v_{rec}; v) f(v)_{MC} dv, \tag{9}$$

where the response function $R(v_{rec}, v)$ is the same as appeared in Equation 5. As was seen in Section 6.1, the unfolding is, by construction, insensitive to details of the prior fragmentation function $f(v)_{MC}$. This is however not necessarily so for the case of the response function. In fact, the response function derived from the Monte Carlo is assumed to be correct and so it is crucial that $R(v_{rec}; v)$ be as close to the situation in the data as possible. A number of corrections were therefore applied to the Monte Carlo to account for known discrepencies with the data for quantities sensitive to the reconstructed kinematical variables v:

- The reconstructed energy distributions per charged or neutral particle were separately shifted and smeared [4] in the Monte Carlo to bring them into better agreement with the data (based on a χ^2-histogram comparison).
- The multiplicity of:
  - fragmentation charged particles (identified by a cut on the TrackNet<0.5),
  - b-hadron weak decay products (identified by a cut on the TrackNet>0.5),
  - neutral particles,

  were fixed in the Monte Carlo, by a weighting function, to be that seen in the data.

[4] For charged particles the shift in the mean was 0.01 GeV and a Gaussian smearing of 3% (relative) applied. For neutral clusters the corresponding numbers were 0.04 GeV and 20%.

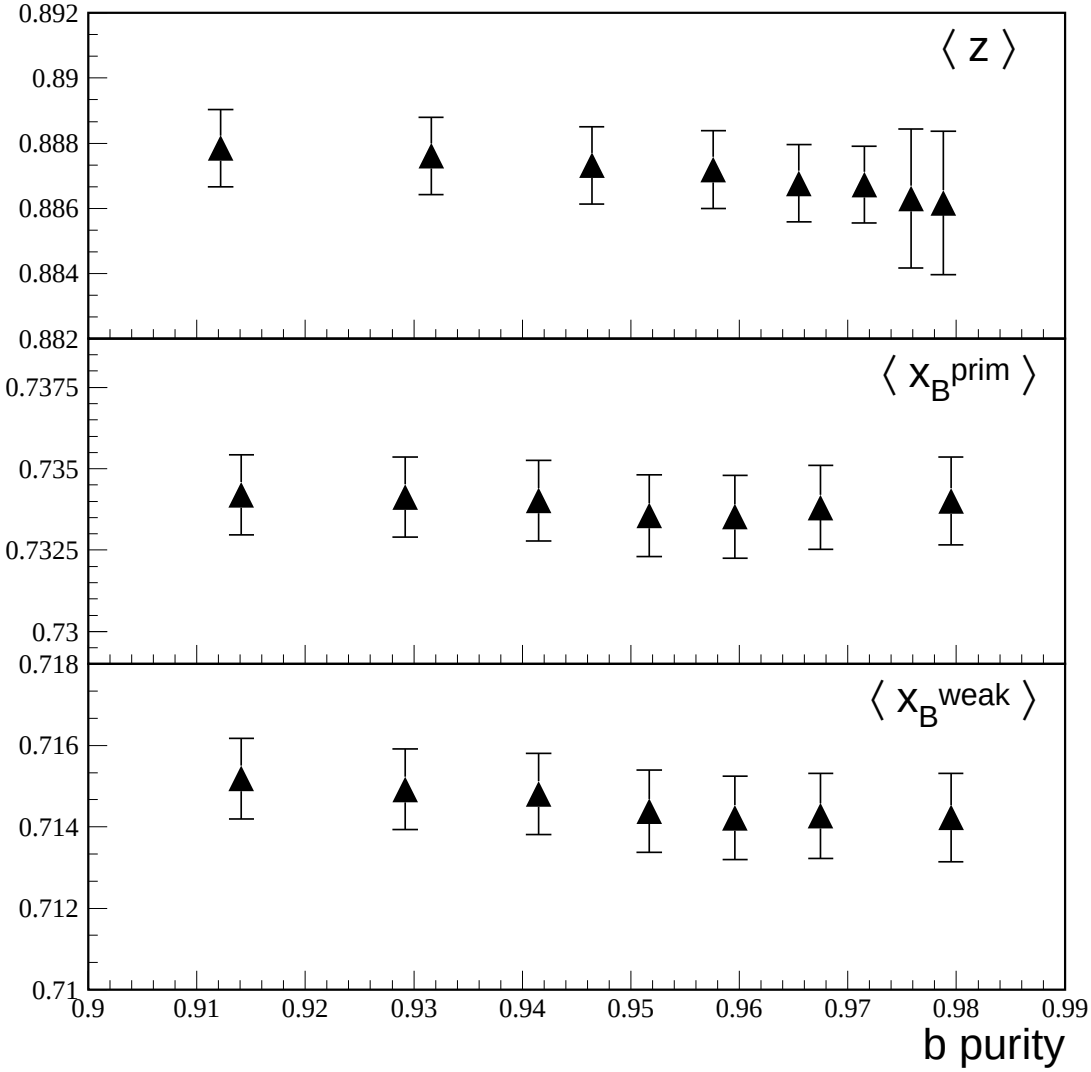

Figure 8: Plots showing $\left\langle x_B^{weak} \right\rangle, \left\langle x_B^{prim} \right\rangle$ and $\langle z \rangle$ as a function of scanning over the analysis cut for b-tagging, represented in the plot as $Z^0 \to b\bar{b}$ purity.

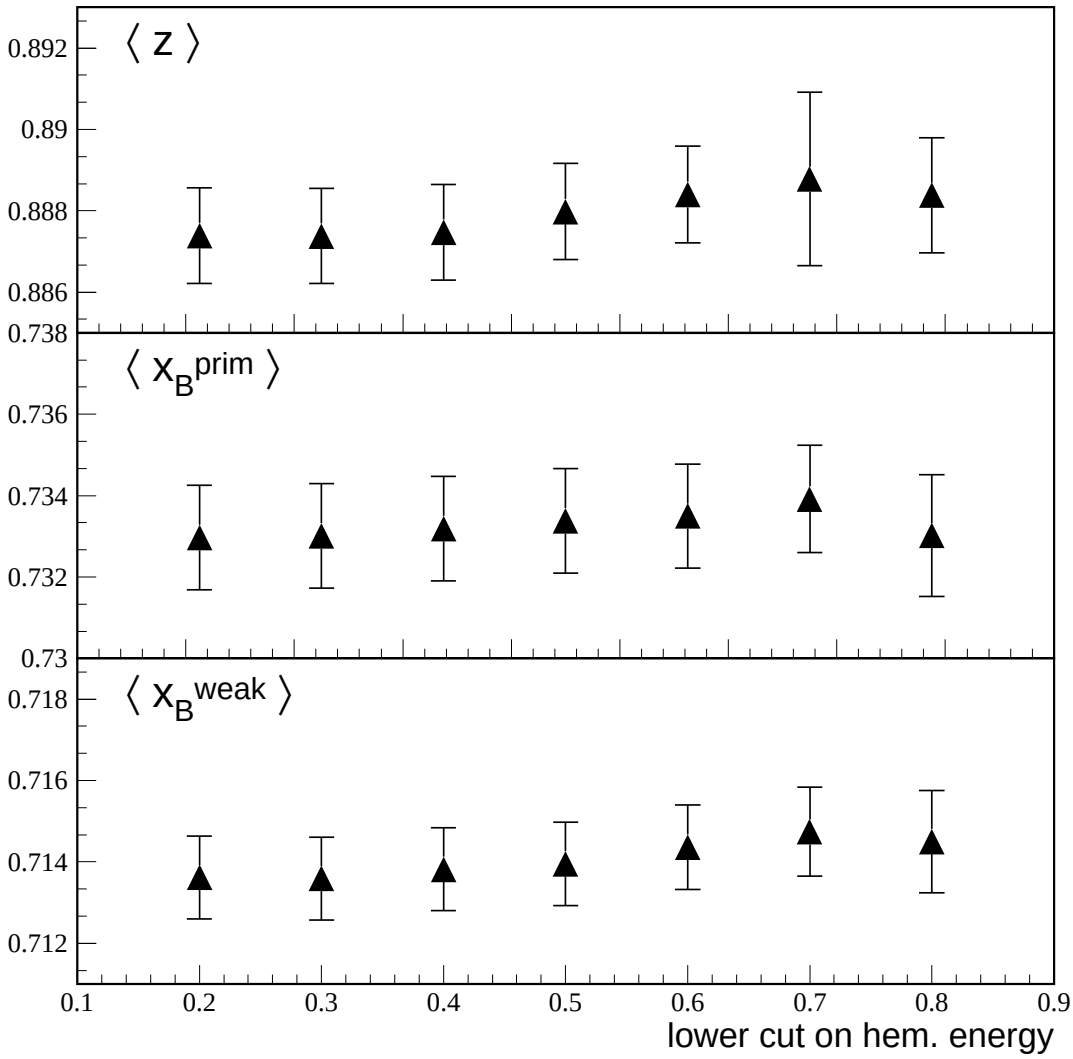

Figure 9: Plots showing $\left\langle x_B^{weak} \right\rangle, \left\langle x_B^{prim} \right\rangle$ and $\langle z \rangle$ as a function of scanning over the analysis cut for $x_{hem} = E_{hem}/E_{beam}$.

- After applying the above two corrections, any residual difference between data and Monte Carlo in the total hemisphere energy (neutral and charged) was accounted for by a further weigthing function.

To account for any systematic uncertainty, due to our imperfect knowledge of energy reconstruction, an error was assigned equal to the full difference seen compared to the default results when the hemisphere energy weight was on/off. In order to take a conservative approach to the energy systematic evaluation, separate errors were also assigned for the particle energy shifting/smearing and multiplicity corrections. The change in the results were taken when the shifting/smearing procedure was turned off and when the multiplicity weights were varied between 50% and 150% of their nominal values. These error contributions can be found listed in Table 3.

A further crosscheck was made for the case of x_B^{weak} by using a different choice for E_B^{weak} other than the Bayesian neural network variable described in Section 4. For this test, E_B^{weak} was reconstructed by the rapidity algorithm mentioned in Section 4 and corrected for missing neutral energy by a parameterisation of e.g. the hemisphere energy as found in the Monte Carlo. Repeating the analysis, the change seen in the result for x_B^{weak} was -0.0011 and is well contained within the assigned total systematic.

7.4 Monte Carlo Weights

The remaining systematic contributions concern quantities that the simulation was weighted for by default, in order to account for known discrepancies with the data. Weights were constructed to change the lifetimes and production fractions of the b-hadron species to more recent world average values:

$$\begin{array}{ll} \tau(B^+) = (1.647 \pm 0.016)\text{ps}, & f(B^+) = (39.9 \pm 1.1)\% \\ \tau(B_d^0) = (1.546 \pm 0.018)\text{ps}, & f(B_d^0) = (39.9 \pm 1.1)\% \\ \tau(B_s^0) = (1.464 \pm 0.06)\text{ps}, & f(B_s^0) = (9.9 \pm 1.1)\% \\ \tau(b-baryon) = (1.208 \pm 0.05)\text{ps}, & f(b-baryon) = (10.3 \pm 1.8)\% \end{array}$$

Systematic uncertainties from these sources were based on varying them within the quoted one standard deviation errors for the case of the lifetimes and by switching the weights on/off for the case of the production fractions. The 'hemisphere quality' was a quantity flagging the presence of potentially badly reconstructed tracks in the hemisphere. Improved agreement with the data was achieved in many reconstructed quantities by weighting the hemisphere quality distribution in Monte Carlo to agree with that seen in data. The change induced by varying the weight between 50% and 150% of the nominal value was assigned as a systematic error.

By default the production rate of excited B^{**} states was adjusted in the simulation to be 30% with respect to B meson hemispheres. This rate was then varied by $\pm 6\%$ and half the change seen in the results assigned as a systematic. In addition, the reconstruction of E_B^{weak} is sensitive to the energy of the primary decay track in B^{**} decay e.g. the energy of the pion in $B^{0**} \to B^+\pi^-$. This has the consequence that the higher the Q-value for B^{**} decay, the lower is the energy of the weakly decaying b-hadron state. Recent studies of B^{**} rates [12] suggests a mean Q-value higher by $\sim$ 300 MeV than the default value assumed in the Monte Carlo which corresponds to $\sim$ 300 MeV shift lower in E_B^{weak}. This

shift was applied in the Monte Carlo for cases were a B^{**} state was present and the change in the result for x_B^{weak} assigned as a systematic error.

Systematic errors from the B^* rate, K_s^0 rate and the b-hadron semi-leptonic branching ratio were accounted for by changing their values in the Monte Carlo by the same relative error quoted on current world averages [13]. In addition an error was assigned due to changing the 'wrong-sign' D_s production rate i.e. D_s production from $W^{\pm}$ decay by 100%.

Finally a weight was applied to the Monte Carlo based on the results of a double hemisphere tagging analysis [14] to evaluate the efficiency to tag $Z^0 \to c\bar{c}$ events and $Z^0 \to b\bar{b}$ events as a function of the standard DELPHI b-tag. At the analysis working point of $b\bar{b}$ purity of 96%, the correction to the b efficiency was only of the order 1% but the correction to the c efficiency was of order 20%. A systematic from this source was conservatively applied as being the full difference in the results when this weight was removed.

7.5 Generator Parameter Dependence

The fact that z was not directly measurable in the experiment and had to be indirectly unfolded from other correlated quantities, meant that the unfolding of z was a more difficult task than for x_B. In the case of z, the unfolding is a two-step process where not only must z be unfolded from the measured z_{rec}, but also the mapping of z from the correlated quantities used to reconstruct it, has to be unravelled. The fact that z is defined only at the start of the non-perturbative phase makes it dependent on the details of the point at which perturbative hadronisation ends and the non perturbative phase starts. An investigation was therefore carried out in order to evaluate the form of this dependence.

The most relevant parameters of the JETSET 7.3 parton shower generator to consider are Λ_{QCD} and Q_0 (the energy scale beyond which non-perturbative modeling begins). At the level of the generator, z is independent of both Λ_{QCD} and Q_0 but this is not true of the reconstructed quantity, z_{rec}. As described in Section 5, z_{rec} is unfolded via a neural network from inputs that are essentially various estimators of x. If x is a function of either Λ_{QCD} or Q_0 therefore, the response function $R(v_{rec}; v)$ would also change, via Equation 9, which in turn would lead to a dependence in the unfolded result $\langle z \rangle$.

The approach used was therefore to take the observed dependence of $\langle x_B^{weak} \rangle$ as a function of Λ_{QCD} and Q_0 (Figure 10(a) and (b)) and then to translate this into the expected variation in $\langle z \rangle$ by using the observed mapping between $\langle x_B^{weak} \rangle$ and $\langle z \rangle$ given by the generator (Figure 10(c)). As seen in Figure 10(a) the dependence on Q_0 was found to be essentially flat and so was not considered further, whereas the variation with Λ_{QCD} was parameterised as, $\langle x_B^{weak} \rangle = -0.12(\pm 0.02)\, \Lambda_{QCD}\ /\mathrm{GeV} + 0.73(\pm 0.01)$. This result was then combined with a parameterisation of the mapping in Figure 10(c), $\langle x_B^{weak} \rangle = 0.63(\pm 0.01)\, \langle z \rangle + 0.15(\pm 0.01)$, to give the required dependence of z as a function of Λ_{QCD},

$$\langle z \rangle \;=\; 0.8872 - 0.19(\pm 0.01)\, (\Lambda_{QCD}\ /\mathrm{GeV} - 0.346)$$

The errors quoted are statistical originating from the parameter fits illustrated in Figure 10. Note that for a value of $\Lambda_{QCD} = 0.346$ i.e. the value set in our default Monte Carlo sample, we recover our previous result quoted in Section 6.2 as expected.

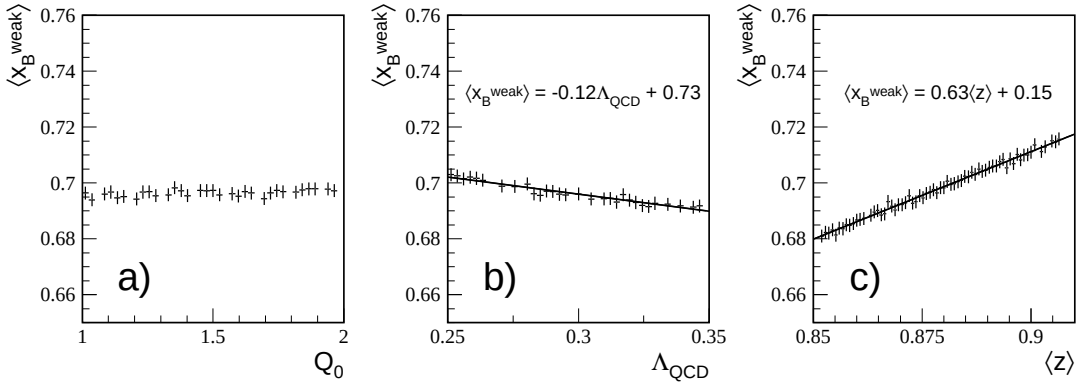

Figure 10: The dependence of $\langle x_B^{weak} \rangle$ at the Monte Carlo generator level, as a function of (a)Q_0 and (b)Λ_{QCD} and (c) $\langle z \rangle$.

8 Fits to Hadronisation Models

The unfolded distributions $f(x_B^{prim}), f(x_B^{weak})$ and $f(z)$ have been compared with functional forms in common use (listed in Table 4) that are implemented inside Monte Carlo generators as functions of z.

In the case of $f(x_B)$, the Lund, Lund-Bowler and Peterson models were compared to the data by generating many Monte Carlo samples spanning a large range of function parameter values, and constructing a binned χ^2 value with the data. Details of the Monte Carlo generator used are listed below:

Monte Carlo Generator	JETSET/PYTHIA 6.156
Perturbative ansatz	Parton shower ($\Lambda_{QCD} = 0.297$ GeV,$Q_0 = 1.56$ GeV)
Non-perturbative ansatz	String fragmentation
Fragmentation function	according to tested model

In the case of $f(z)$, it was possible to fit directly the models developed by Peterson et al., Kartvelishvili et al. and Collins-Spiller to the unfolded data points. The Lund and Lund-Bowler models were compared to the data in the same way as for $f(x_B)$.

In all cases the χ^2 was calculated using the full covariance matrix for the unfolded points. The best fits to the data for $f(x_B^{prim}), f(x_B^{weak})$ and $f(z)$ are shown in Figure 11 and the resulting parameter values are listed in Table 5. The data suggests that the Lund and Lund-Bowler functions are better fits than those explicitly constructed to describe the fragmentation of heavy quarks e.g. the Peterson et al. function.

error source	$\Delta\langle x_B^{prim}\rangle$	$\Delta\langle x_B^{weak}\rangle$	$\Delta\langle z\rangle$
number of degrees of freedom	0.0003	0.0003	0.0003
number of knots	0.0002	0.0003	0.0002
neutral energy smearing	0.0030	0.0026	0.0023
fragmentation track multiplicity	0.0033	0.0031	0.0020
b-decay track multiplicity	0.0004	0.0004	0.0004
neutral multiplicity	0.0011	0.0010	0.0003
hemisphere energy x_{hem}	0.0002	0.0003	0.0008
K^0 rate	0.0005	0.0005	0.0006
B^* rate	0.0009	0.0006	0.0002
semileptonic decay rate	0.0001	0.0001	0.0002
wrong sign charm rate	0.0001	0.0001	0.0002
$g \to b\bar{b}$	0.0008	0.0007	0.0001
c and b-quark efficiency	0.0001	0.0001	0.0001
hemisphere quality	0.0020	0.0018	0.0017
B^{**} rate	0.0018	0.0011	0.0038
B^{**} Q-value dependence	—	-0.0019	—
B lifetimes	0.0004	0.0004	0.0003
B production fractions	0.0004	0.0002	0.0009
total systematic error	0.0055	$^{+0.0049}_{-0.0052}$	0.0054

Table 3: Systematic uncertainties on the mean values of the unfolded distributions in x_B^{prim}, x_B^{weak} and z. The total is the sum in quadrature of all contributions.

	Fitting to $f(x_B^{prim})$		Fitting to $f(x_B^{weak})$		Fitting to $f(z)$	
Model	Parameters	χ^2/ndf	Parameters	χ^2/ndf	Parameters	χ^2/ndf
Kart.	—	—	—	—	$\alpha_b = 14.6 \pm 0.7$	36/3
CS	—	—	—	—	$\epsilon_b = 0.001\pm < 0.001$	536/3
Peterson	$\epsilon_b = 0.00286$	245/9	$\epsilon_b = 0.00286$	287/9	$\epsilon_b = 0.002\pm < 0.001$	187/3
Lund	$a = 2.1$ $b = 0.9$	53/8	$a = 0.9$ $b = 0.5$	42/8	$a = 2.4$ $b = 0.7$	2/2
Lund-Bowler	$a = 1.4$ $b = 2.7$	43/8	$a = 1.0$ $b = 2.3$	35/8	$a = 2.04$ $b = 2.48$	1/2

Table 5: Results of the $f(x_B)$ and $f(z)$ hadronisation model fits. The fits to $f(x_B)$ were made in the range $[0.1, 1.0]$ and to $f(z)$ in the range $[0.71, 1.0]$. Note that the fits are made to data points with statistical and systematic error assignments.

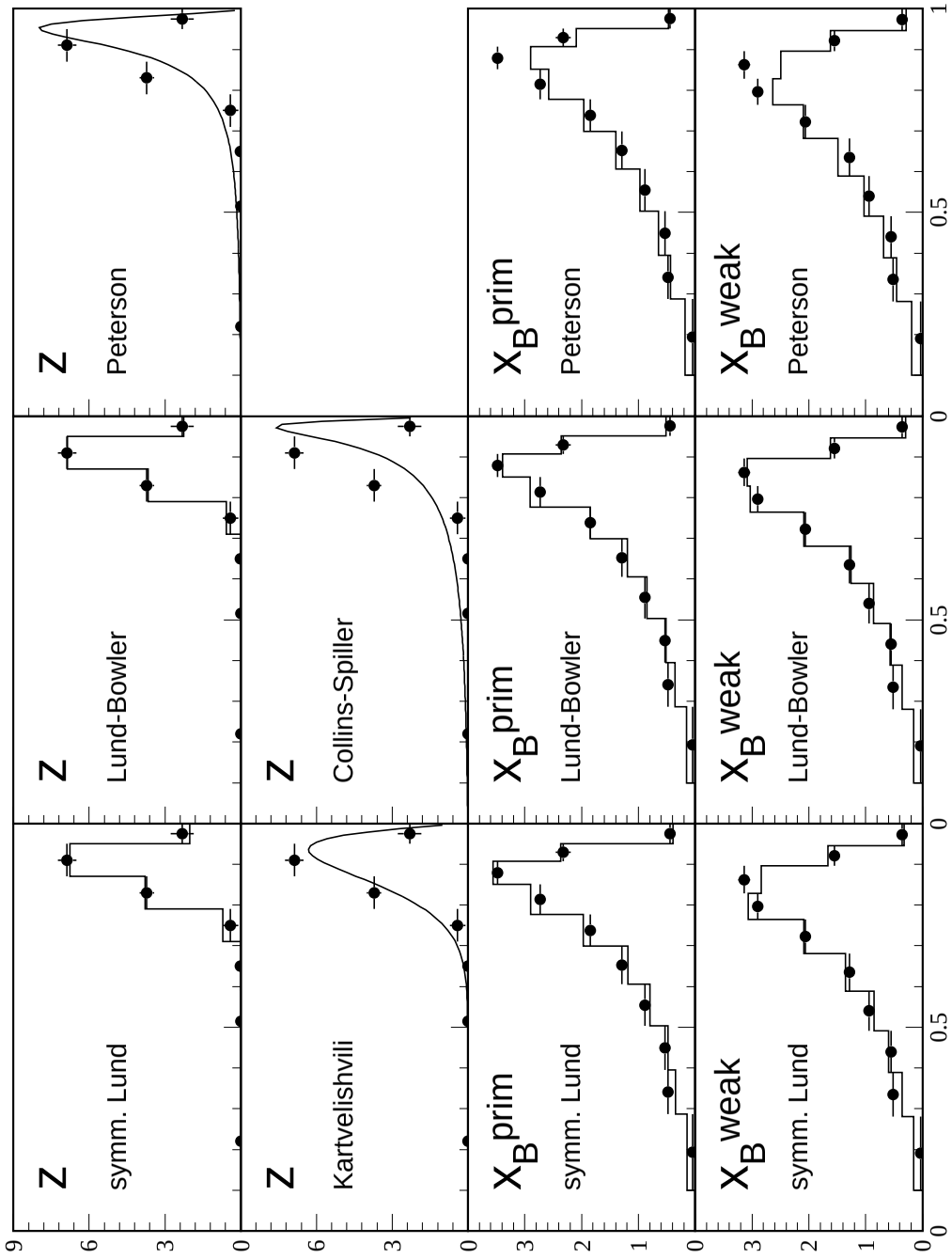

Figure 11: The result of fitting various hadronisation model functions to the unfolded $f(x_B^{prim})$, $f(x_B^{weak})$ and z distributions.

	$f(z)$
Kart	$z^{\alpha_b}(1-z)$
CS	$\left(\frac{1-z}{z}+\frac{(2-z)\epsilon_b}{1-z}\right)(1+z^2)\left(1-\frac{1}{z}-\frac{\epsilon_b}{1-z}\right)^{-2}$
Peterson	$\frac{1}{z}\left(1-\frac{1}{z}-\frac{\epsilon_b}{1-z}\right)^{-2}$
Lund	$\frac{1}{z}z^{a_\alpha}\left(\frac{1-z}{z}\right)^{a_\beta}\exp\left(-\frac{bm_\perp^2}{z}\right)$
Lund - Bowler	$\frac{1}{z^{1+r_Q bm_Q^2}}z^{a_\alpha}\left(\frac{1-z}{z}\right)^{a_\beta}\exp\left(-\frac{bm_\perp^2}{z}\right)$

Table 4: Various hadronisation functions tested to describe the unfolded data distributions: **Kart**: V.G. Kartvelishvili et al.[15], **CS**: P. Collins, T. Spiller[16], **Peterson**: C. Peterson et al.[7], **Lund**: B. Andersson et al.[18], **Lund-Bowler**: M. Bowler[19]. Note that the Lund - Bowler model is implemented in the generator in an extended version according to [20].

9 QCD analysis of the measured b-fragmentation distribution

In section 8, parameters for the non-perturbative b-fragmentation component which has to be used inside the Pythia 6.156 parton Monte Carlo generator, have been determined in such a way that the resulting x_B^{weak} distribution agrees with the measurements. These parameters have been obtained for several models.

Another approach consists in determining the non-perturbative component which has to be used in conjunction with a perturbative QCD evaluation of the b-fragmentation distribution. Parameters of this non-perturbative component depend on the values of the parameters and on the order used in the perturbative QCD computation. According to factorization, the non-perturbative component extracted in this way is expected to be applicable in other processes than $e^+e^- \to b\overline{b}$ as long as the same approximations and the same values for the perturbative QCD part have been used.

In the following we indicate the main lines of the perturbative QCD approach and extract values for the parameters of the corresponding non-perturbative component.

9.1 The perturbative QCD component of the b-fragmentation distribution

The perturbative QCD fragmentation function is evaluated according to the approach presented in [23]. This next to leading log. (NLL) accuracy calculation for the inclusive b-quark production cross section in e^+e^- annihilation, generalizes previous calculations by resumming the contribution from soft gluon radiation to all perturbative orders and to NLL accuracy. These contributions play an important role at large x.

These computations are done using Mellin transformed distributions:

$$\tilde{F}(N, Q^2) = \int_0^1 dx\ x^{N-1}\ F(x, Q^2) \tag{10}$$

where N is complex. When N is a positive integer, the Mellin transformation corresponds to the distribution of algebraic moments of order N-1.

In the Mellin conjugate space, named also the moment's space in the following, the perturbative fragmentation function ($\tilde{D}_{pert.}$) is simply the product of two quantities:

$$\tilde{D}_{pert.}(N) = \tilde{C}_{e^+e^-}(N) \times \tilde{D}_{parton}(N) \tag{11}$$

The coefficient function, $\tilde{C}_{e^+e^-}(N)$, determines the physics of the hard process occuring at the scale (Q) which corresponds to the center of mass energy, in e^+e^- collisions. Roughly, it is the $b\overline{b}$ pair creation cross section. This quantity is process dependent. On the other hand, the factor $\tilde{D}_{parton}(N)$ is not process dependent. It reflects the physics of the evolution of parton distributions from the initial to the scale of the heavy quark's mass (typically the gluon radiation). The factorization scheme adopted in the calculation [23] is $\overline{\text{MS}}$.

Soft gluons radiation contributes to the logarithm of the fragmentation function large logarithmic terms of the type $\alpha_s^n(logN)^p$, with $p \leq n+1$. These terms appear in all the perturbative orders n in α_s. In the NLL accuracy calculation [23] the two largest terms, corresponding to $p = n+1$ and n, have been resummed at all perturbative orders. The

calculation is expected to be reliable when N is not too large (typically less than 20). To obtain the distribution for the variable x, from the results in the moment's space, one should apply the inverse Mellin transformation. This can be done by integrating over a contour in N. When x gets closer to 1, large values of N contribute and thus the perturbative fragmentation distribution is not reliable and becomes even un-physical in these regions. This behaviour affects also values of the distribution at lower x as moments of this distribution are fixed.

Uncertainties on the moments of the perturbative QCD distribution have been evaluated by varying, between $Q/2$ and $2Q$, the values of the scales μ and μ_F introduced in the calculation. In [23] it has been assumed that $\mu = \mu_F$. Corresponding variations induced on the moments of the perturbative component are displayed in Figure 12.

9.2 Measurement of the moments of the non-perturbative QCD component of the b-fragmentation distribution

The variation of the measured b-fragmentation distribution, in the moment's space, is displayed in Figure 12. In Tables 6 and 7, are given respectively the values for the first five moments and the corresponding full error matrix, taking into account the different sources of correlations.

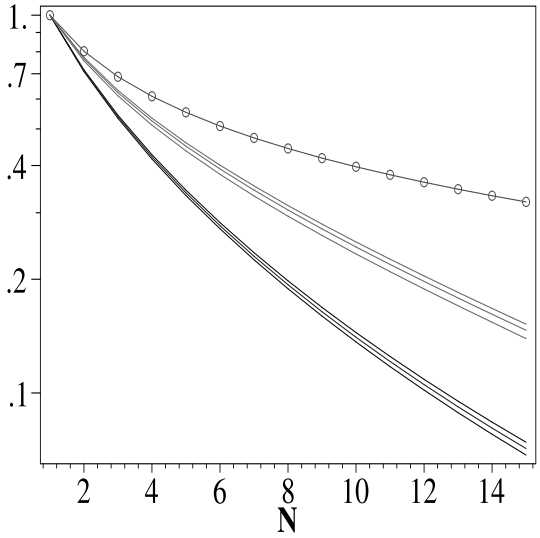

Figure 12: *Moments of the measured (full lines) x_B^{weak} distribution, of the perturbative QCD component (dashed lines) and of the generated distribution obtained in JETSET before hadronization (line with circles).*

On Figure 12, moments of the perturbative QCD component have been also displayed and corresponding values are given in Table 6 for the first five moments. From these two distributions, it is possible to extract the moments of the non-perturbative QCD component which has to be introduced to account for the measured distribution. In the moment's space this distribution is obtained simply by dividing the two previous distributions:

$$\tilde{D}_{non-pert.}(N) = \frac{\tilde{D}_{meas.}(N)}{\tilde{D}_{pert.}(N)} \tag{12}$$

Component	$< x >$	$< x^2 >$	$< x^3 >$	$< x^4 >$	$< x^5 >$
Measured	0.7153	0.5401	0.4236	0.3406	0.2789
Pert. QCD	0.7666	0.6239	0.5246	0.4502	0.3917
Non Pert. QCD	0.9331	0.8657	0.8075	0.7566	0.7120

Table 6: *Values for the moments of the measured x_B^{weak} distribution and of the perturbative QCD component. Moments quoted in the last line for the non-perturbative QCD component correspond to the ratio between the numbers given in the two previous lines.*

2.7297×10^{-5}	3.3019×10^{-5}	3.3395×10^{-5}	3.2125×10^{-5}	3.0275×10^{-5}
3.3019×10^{-5}	4.0560×10^{-5}	4.1427×10^{-5}	4.0107×10^{-5}	3.7963×10^{-5}
3.3395×10^{-5}	4.1427×10^{-5}	4.2589×10^{-5}	4.1412×10^{-5}	3.9320×10^{-5}
3.2125×10^{-5}	4.0107×10^{-5}	4.1412×10^{-5}	4.0389×10^{-5}	3.8437×10^{-5}
3.0275×10^{-5}	3.7963×10^{-5}	3.9320×10^{-5}	3.8437×10^{-5}	3.6642×10^{-5}

Table 7: *Values of the errors on measured moments of the x_B^{weak} distribution.*

In Figure 12 have been also reported the moments of the distribution obtained at the level of the JETSET parton-shower simulation, just before the hadronization step. It can be seen that the analytic computation accounts for much softer gluon radiation than given by the present simulation.

In Table 8, the fitted values of the parameters corresponding to a few models, and using the first five moments in the fit, have been given.

Model	Fitted parameters	χ^2/NDF
Kartvelishvili et al.	17.07 ± 0.57	115/4
Peterson	$(1.55 \pm 0.13)10^{-3}$	189/4
Collins Spiller	$(4.6 \pm 1.2)10^{-5}$	845/4

Table 8: *Values for the fitted parameters of the non-perturbative QCD component.*

None of the models, considered in Table 8, provides a good description of the data. This is due to the accuracy of present measurements. Values of fitted parameters, given in table 8, differ from those obtained in table 5 as the perturbative QCD components used in the two cases are different.

9.3 Measurement of the x-dependence of the non-perturbative QCD component of the b-fragmentation distribution

Present measurements can be used to extract also the x-dependence of the non-perturbative component without the need to use any specific model. It is necessary to apply an inverse Mellin transformation to the distribution of the moments of the non-perturbative component which was given by Equation 12. This can be done if the measured and the perturbative component are determined for all values of N (N being a complex variable). The N-dependence of the perturbative component can be found in [23].

To determine the N-dependence of the measured fragmentation distribution, this distribution was firstly fitted as a function of x using the following expression which depends on five parameters and gives a good description of the measurements.

$$D(x) = p_0 \times (p_1 x^{p_2}(1-x)^{p3} + (1-p_1)x^{p_4}(1-x)^{p5}) \tag{13}$$

where p_0 is a normalisation coefficient. The Mellin transformation of $D(x)$ gives the value of the measured fragmentation distribution for all specified value of N. Applying an inverse Mellin transformation to $\tilde{D}_{non-pert.}(N)$ deduced from Equation 12, the x-dependence of the non-perturbative component is obtained. It has been displayed in Figure 13. It shows an unphysical behaviour in some x-intervals which compensate the corresponding unphysical properties of the perturbative component.

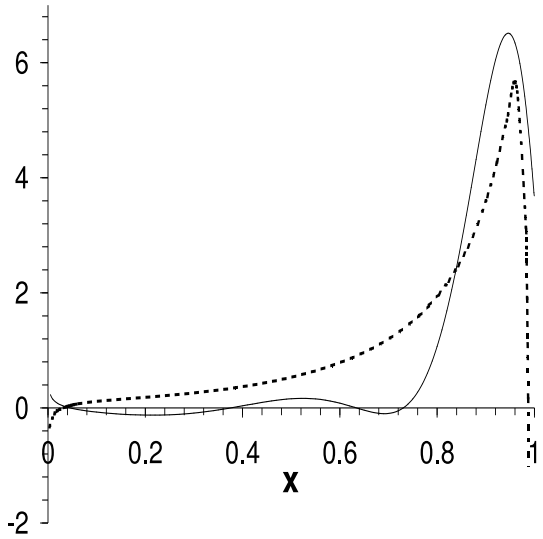

Figure 13: *x-dependence of the perturbative (dotted line) and non-perturbative (full line) QCD components of the measured b-fragmentation distribution.*

The high-x ($x > 0.96$) behaviour of this distribution has been studied. This region corresponds to high-N values where the perturbative approach fails. As a result, the high-x behaviour of the perturbative and non-perturbative QCD components is non-physical; the distributions oscillate. To have a numerical control of these distributions, in these regions, we have chosen to do the evaluation up to a given maximum value, x_{max}, above which the distributions are assumed to be equal to zero. Moments of these truncated distributions show a small discrepancy when compared with the moments of the full distributions. To correct for this effect, x_{max} is chosen such that the difference between moments is a constant value and a Dirac distribution, corresponding to this difference, is added at $x = 1$. A typical value for x_{max} is 0.997 and the Dirac component corresponds to 5% of the distribution.

It has been verified that the convolution of the perturbative and non-perturbative components reproduce the measured b-fragmentation distribution.

10 Conclusion

Using the DELPHI 1994 Z^0 statistics the b-quark fragmentation functions $f(x_B^{prim})$, $f(x_B^{weak})$ and, for the first time, $f(z)$ have been extracted from the data.

An unfolding method based on regularisation techniques was shown to be robust against details of the prior fragmentation function used to generate the Monte Carlo sample and functioned correctly even when the measured distribution was only weakly correlated to the truth. The results and covariance matrices are detailed in Appendix A. In terms of the first two central moments of the unfolded functions, the results obtained were:

	Mean	Variance
$f(x_B^{prim})$	$0.7346 \pm 0.0008(stat.) \pm 0.0055(syst.)$	$0.0318 \pm 0.0005(stat.)$
$f(x_B^{weak})$	$0.7153 \pm 0.0007(stat.)^{+0.0049}_{-0.0052}(syst.)$	$0.0300 \pm 0.0004(stat.)$
$f(z)$	$0.8872 \pm 0.0012(stat.) \pm 0.0054(syst.)$	$0.0027 \pm 0.0001(stat.)$

Figure 14 presents a comparison of the results from this analysis for x_B^{weak} compared to two other recent analyses employing an unfolding technique, from ALEPH[21] and SLD[22]. The unfolded fragmentation functions are seen to be quite consistent in shape within the quoted errors. It should be noted that the ALEPH and SLD unfolded points, in contrast to the present analysis, are almost fully correlated bin-to-bin.

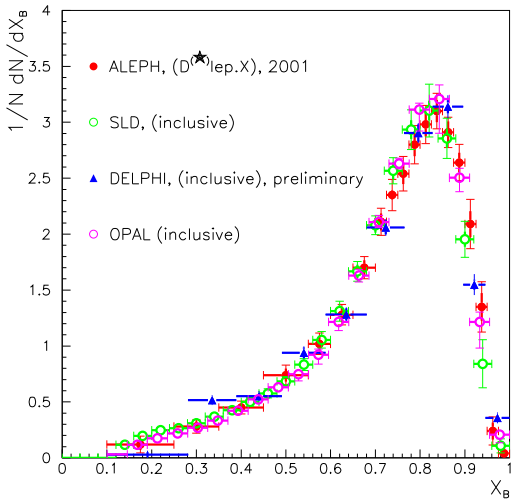

Figure 14: Unfolded $f(x_B^{weak})$ distributions from ALEPH, OPAL [24], SLD and the current analysis.

The dependence of the result for z as a function of the Monte Carlo generator parameters Q_0 and Λ_{QCD} was also studied. The variation with Q_0 was found to be negligible and as a function of Λ_{QCD} the following relationship was measured,

$$\langle z \rangle = 0.8872 - 0.19(\pm 0.01)\,(\Lambda_{QCD}\ /\mathrm{GeV} - 0.346).$$

It should be emphasized that the results for the case of $f(z)$ are model dependent and should be evaluated always within the framework of the Monte Carlo generator used, in this case JETSET 7.3 parton shower.

The variation in the results as a function of the event thrust gives an important insight into the effect of perturbative QCD effects on the hadronisation process. As expected, $\langle x_B \rangle$ is seen to rise with increasing thrust, whereas $\langle z \rangle$ is approximately constant. Fits to hadronisation model functions show, particularly for the case of z unfolding, that many of the most commonly used parameterisations are poor representations of the data.

In addition, the measured b-quark fragmentation distribution has been analysed using recent perturbative QCD computations. It was found that the models developed by Peterson et al., Kartvelishvili et al. and Collins/Spiller for the non-perturbative component are unable to reproduce the measurements when used in conjunction with a NLL perturbative calculation. Using present measurements the non-perturbative component has been extracted in the space of moments, independently of any model assumption, and its x-dependence has been obtained. This extracted non-perturbative QCD distribution is also expected to be process independent and could be used in another environment as long as the same NLL perturbative QCD approach is followed in that new context.

Appendix A: Unfolded Points and Covariance Matrices

bin borders	value	stat. error	syst. error	$\sqrt{\sigma^2_{stat} + \sigma^2_{syst}}$
0.100 – 0.287	0.051	0.025	0.018	0.031
0.287 – 0.395	0.481	0.032	0.022	0.039
0.395 – 0.503	0.536	0.038	0.024	0.045
0.503 – 0.606	0.887	0.049	0.033	0.059
0.606 – 0.699	1.294	0.059	0.029	0.066
0.699 – 0.777	1.849	0.069	0.022	0.072
0.777 – 0.851	2.728	0.067	0.036	0.076
0.851 – 0.907	3.478	0.066	0.049	0.082
0.907 – 0.951	2.324	0.051	0.122	0.133
0.951 – 1.000	0.442	0.045	0.061	0.076

Table 9: The unfolding result, per bin, for $f(x_B^{prim})$.

bin	1	2	3	4	5	6	7	8	9	10
1	0.628									
2	-0.560	1.032								
3	0.111	-0.709	1.454							
4	-0.055	0.351	-1.389	2.412						
5	-0.035	-0.009	0.754	-2.209	3.455					
6	0.063	-0.152	-0.195	1.233	-3.072	4.717				
7	-0.066	0.165	-0.068	-0.363	1.479	-3.538	4.489			
8	0.039	-0.105	0.101	0.038	-0.544	1.843	-3.318	4.310		
9	-0.015	0.052	-0.102	0.145	-0.081	-0.321	1.025	-2.260	2.612	
10	0.004	-0.013	0.008	0.029	-0.139	0.411	-0.753	1.329	-1.654	2.062

Table 10: The statistical covariance matrix, in units of 10^{-3}, for the unfolded bins in $f(x_B^{prim})$.

bin	1	2	3	4	5	6	7	8	9	10
1	0.959									
2	-0.729	1.502								
3	0.303	-0.519	2.022							
4	0.180	0.565	-1.094	3.488						
5	0.175	0.219	1.226	-1.948	4.310					
6	0.206	-0.310	-0.358	1.329	-3.054	5.202				
7	0.134	0.060	0.027	-0.267	1.904	-3.111	5.798			
8	-0.395	-0.498	-0.611	-1.075	-1.443	1.753	-4.184	6.711		
9	-1.289	-0.593	-1.919	-2.439	-2.824	-1.359	-1.406	1.973	17.578	
10	-0.831	-0.179	-1.000	-1.037	-1.484	-0.030	-1.876	3.667	4.575	5.800

Table 11: The total (i.e. including statistical and systematic errors) covariance matrix, in units of 10^{-3}, for the unfolded bins in $f(x_B^{prim})$.

bin borders	value	stat. error	syst. error	$\sqrt{\sigma^2_{stat} + \sigma^2_{syst}}$
0.100 – 0.281	0.029	0.022	0.029	0.036
0.281 – 0.389	0.516	0.030	0.032	0.044
0.389 – 0.491	0.552	0.037	0.028	0.046
0.491 – 0.589	0.940	0.046	0.031	0.056
0.589 – 0.681	1.282	0.054	0.033	0.064
0.681 – 0.764	2.060	0.063	0.040	0.075
0.764 – 0.828	2.904	0.077	0.043	0.088
0.828 – 0.896	3.139	0.059	0.084	0.103
0.896 – 0.946	1.549	0.044	0.083	0.094
0.946 – 1.000	0.358	0.038	0.030	0.048

Table 12: The unfolding result, per bin, for $f(x_B^{weak})$.

bin	1	2	3	4	5	6	7	8	9	10
1	0.482									
2	-0.463	0.924								
3	0.139	-0.664	1.373							
4	-0.029	0.205	-1.201	2.159						
5	-0.048	0.116	0.526	-1.883	2.950					
6	0.040	-0.188	-0.092	1.090	-2.658	3.985				
7	-0.046	0.212	-0.146	-0.434	1.691	-3.664	5.886			
8	0.025	-0.117	0.180	-0.081	-0.322	1.189	-3.325	3.483		
9	-0.008	0.043	-0.088	0.099	-0.017	-0.220	1.205	-1.921	1.966	
10	0.001	-0.008	0.004	0.028	-0.104	0.262	-0.714	0.940	-1.141	1.453

Table 13: The statistical covariance matrix, in units of 10^{-3}, for the unfolded bins in $f(x_B^{weak})$.

bin	1	2	3	4	5	6	7	8	9	10
1	1.301									
2	-1.173	1.943								
3	0.586	-0.884	2.133							
4	0.387	0.221	-0.835	3.105						
5	0.160	-0.034	0.991	-1.861	4.040					
6	0.511	-0.435	0.021	1.953	-2.922	5.619				
7	0.241	-0.198	0.045	-0.592	2.844	-3.837	7.754			
8	-1.489	0.530	-1.286	-1.848	-1.974	-0.441	-5.329	10.509		
9	-1.602	0.830	-1.730	-1.667	-1.559	-2.162	-0.126	3.931	8.910	
10	-0.460	0.036	-0.570	-0.695	-0.530	-0.262	-0.947	2.792	1.011	2.343

Table 14: The total (i.e. including statistical and systematic errors) covariance matrix, in units of 10^{-3}, for the unfolded bins in $f(x_B^{weak})$.

bin borders	value	stat. error	syst. error	$\sqrt{\sigma^2_{stat} + \sigma^2_{syst}}$
0.000 – 0.440	0.000	0.000	0.006	0.006
0.440 – 0.590	0.000	0.000	0.000	0.000
0.590 – 0.710	0.005	0.001	0.021	0.021
0.710 – 0.790	0.408	0.050	0.305	0.309
0.790 – 0.870	3.720	0.059	0.281	0.287
0.870 – 0.950	6.875	0.078	0.353	0.362
0.950 – 1.000	2.312	0.154	0.431	0.459

Table 15: The unfolding result, per bin, for $f(z)$.

bin	1	2	3	4	5	6	7
1	0.000						
2	0.000	0.000					
3	0.000	0.000	0.001				
4	0.000	0.000	0.054	2.528			
5	0.000	0.000	-0.038	-1.670	3.474		
6	0.001	0.000	-0.036	-1.824	-1.403	6.120	
7	0.002	0.000	-0.092	-4.100	7.088	-4.543	24.328

Table 16: The statistical covariance matrix, in units of 10^{-3}, for the unfolded bins in $f(z)$.

bin	1	2	3	4	5	6	7
1	0.042						
2	0.000	0.000					
3	0.041	0.002	0.441				
4	0.233	0.032	5.844	95.249			
5	0.198	0.077	-0.026	0.519	82.500		
6	-0.491	-0.118	-4.802	-81.340	-38.097	130.805	
7	-0.328	0.004	-3.159	-45.065	-65.676	-1.493	210.273

Table 17: The total (i.e. including statistical and systematic errors) covariance matrix, in units of 10^{-3}, for the unfolded bins in $f(z)$.

References

[1] B. Anderson, G. Gustafson, G. Ingelman, T. Sjöstrand, Phys. Rep. **97** (1983) 31.

[2] Z. Albrecht, T. Allmendinger, G. Barker, M. Feindt, C. Haag, M. Moch, *BSAURUS-A Package For Inclusive B-Reconstruction in DELPHI* , hep-ex/0102001.

[3] DELPHI Collaboration, Nucl. Instr. and Meth. **A303**(1991) 233;

[4] DELPHI Collaboration, Nucl. Instr. and Meth. **A378**(1996) 57.

[5] Z. Albrecht, M. Feindt, M. Moch, *MACRIB, high efficiency, high purity hadron identification for DELPHI*, hep-ex/0111081.

[6] T. Sjöstrand, *Comp. Phys. Comm.* **39** (1986) 347;
T. Sjöstrand, M. Bengtsson, *Comp. Phys. Comm.* **43** (1987) 367.

[7] C. Peterson, D. Schlatter, I. Schmitt, P.M. Zerwas, Phys. Rev. **D27** (1983) 105

[8] DELPHI Collaboration, P. Abreu et al., Z. Phys **C73** (1996) 11.

[9] G.V. Borisov, DELPHI note, 94-125 PROG 208, (1994);
G.V. Borisov, DELPHI note, 97-94 PHYS 716, (1997).

[10] V. Blobel, *The RUN manual: Regularized Unfolding for High-Energy Physics* OPAL Technical Note TN361, March 8, 1996.

[11] V. Blobel, *Unfolding Methods In High-Energy Physics Experiments*, DESY 84-118 (1984);
V. Blobel, *Proceedings of the 1984 CERN School of Computing*, CERN 85-02 (1985).

[12] *A study of excited b-hadron states with the DELPHI detector at LEP1*, Z. Albrecht et al., submitted to ICHEP 2002 Amsterdam, DELPHI 2002-079 CONF 613.

[13] K. Hagiwara et al., Phys. Rev. **D66** (2002) 010001.

[14] T. Scheidle et al., *Measuring the b and c-Quark Efficiency With a Double Tag Method*, DELPHI note in preparation.

[15] V.G. Kartvelishvili, A.K. Likehoded, V.A. Petrov, Phys. Lett. **B78** (1978) 615

[16] P. Collins, T. Spiller, J. Phys. **G11** (1985) 1289.z

[17] ALEPH Collaboration, Phys. Lett. **B357** (1995) 699.

[18] B. Andersson, G. Gustafson and B. Soderberg, Z. Phys. C **20** (1983) 317.

[19] M. G. Bowler, Z. Phys. C **11** (1981) 169.

[20] D. A. Morris, Nucl. Phys. B **313** (1989) 634.

[21] ALEPH Collaboration, *Study of the fragmentation function of b quarks into B mesons at the Z peak*, Phys. Lett. **B512** (2001) 30-48.;

[22] SLD Collaboration, *Precise measurement of the b quark fragmentation function in Z^0 Boson decays*, Phys. Rev. Lett. **84** (2000) 4300.

[23] M. Cacciari and S. Catani, Nucl. Phys. **B617** (2001) 253, hep-ph/0107138.

[24] OPAL Collaboration, G. Abbiendi et al., *Inclusive Analysis of the b Quark Fragmentation Function in Z Decays at LEP*, OPAL PR359 (2002), to be submitted to Eur. Phys. J. C.