PRÄZISIONSMESSUNGEN DER LEBENSDAUERN VON IDENTIFIZIERTEN b-HADRONEN MIT DEM DELPHI-DETEKTOR

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik der Universität (TH)
Karlsruhe

genehmigte

DISSERTATION

von

Dipl.-Phys. Christian Haag
aus Bruchsal

Tag der mündlichen Prüfung: 16.05.2003

Referent: Prof. Dr. M. Feindt, Institut für Experimentelle Kernphysik

Korreferent: Prof. Dr. G. Quast, Institut für Experimentelle Kernphysik

Abstract

In the early 1980's first measurements of the b-hadron lifetime were able to infer that the average b-hadron lifetime was around 1 ps. Since then many lifetime measurements were performed with increasing precision. Due to the development of sophisticated detectors and reconstruction techniques, the lifetimes of the individual b-hadron-species have been resolved by the LEP experiments and SLD. The latest world averages for the individual lifetimes are $\tau_{B^+} = 1.647 \pm 0.016$ ps, $\tau_{B^0} = 1.546 \pm 0.018$ ps, $\tau_{B_s} = 1.464 \pm 0.057$ ps and $\tau_{\Lambda_b} = 1.208 \pm 0.051$ ps, for the ratio $\tau_{B^+}/\tau_{B^0} = 1.068 \pm 0.016$ and for the average b hadron mean lifetime $\tau_b = 1.564 \pm 0.014$ ps.

The vast majority of previous lifetime measurements was statistically limited by using only a few percent of the available b decay channels. Within this analysis, a fully inclusive approach was employed, using the data set of the DELPHI detector taken at the Z^0 pole during the years 1994 and 1995. For this purpose the analysis made extensive use of advanced analysis techniques such as artificial neural networks and exploited the full capabilities of the DELPHI detector.

After all selection cuts the data sample consisted of approximately 420000 event hemispheres. Dedicated b-hadron identification neural networks, momentum and decay length reconstruction algorithms were developed, based on the BSAURUS package [94].

The results were extracted using a χ^2 fit performed on enhanced samples, which were approximately 70% pure in B^+ and B^0 mesons, resulting in the following values for the lifetimes of B^+ and B^0 mesons and their ratio:

$$\begin{aligned} \tau_{B^+} &= 1.625 \pm 0.013 \text{ (stat) } \pm 0.017 \text{ (syst) ps} \\ \tau_{B^0} &= 1.543 \pm 0.020 \text{ (stat) } \pm 0.033 \text{ (syst) ps} \\ \frac{\tau_{B^+}}{\tau_{B^0}} &= 1.051 \pm 0.019 \text{ (stat) } \pm 0.024 \text{ (syst)} \end{aligned}$$

Utilizing the same method on a sample which was not enhanced in B^+ or B^0 mesons, the average b-hadron lifetime was found to be:

$$\tau_b = 1.568 \pm 0.005 \text{ (stat) } \pm 0.009 \text{ (syst) ps}$$

The results are consistent with the current world averages. They represent a significant improvement in precision over previous measurements, the result for the B^+ and B^0 lifetime ratio being the most precise single measurement of this quantity from Z^0 resonance data. The result of the average b-hadron mean lifetime is currently the most precise extraction of this number worldwide.

Contents

List of Figures

List of Tables

Introduction

> A beginning is a very delicate time.
> Princess Irulan

With the discovery of the *bottom* or *beauty* quark in 1977 (see [1, 2]) in the Υ resonance a successful 25 years period of b physics started. b physics provides important tests of the so-called *Standard Model* and measurements of its parameters, especially the parameters of the quark mixing matrix. This matrix connects the mass eigenstates of the quarks to their flavor eigenstates participating in weak interaction processes. This mixing mechanism enables quarks from one generation to interact with quarks from other generations, and so the b quark can decay e.g. via $b \to c$. Due to the fact that this so-called *Cabibbo-Kobayashi-Maskawa* (CKM) matrix contains a complex phase which is the only established source for CP violation, b physics was focused on because CP violation could be a possible explanation for one of the remaining big puzzles of our days: Up to now, there exists no satisfactory and proven explanation for the asymmetry between matter and anti-matter in the universe.

In 1982 the JADE Collaboration at PETRA set an upper limit on the mean b-hadron lifetime of $\langle \tau_B \rangle < 1.4$ ps [3]. Before 1983, the lack of experimental data on the strength of mixing between the first respectively second fermion generation with the third lead to the prediction that the B lifetime should be of the order of 0.1 ps, assuming, that the strength of mixing between the second and third generations was the same as that existing between the first and second. There was, however, evidence from the endpoint of the semileptonic decay spectrum of b-hadrons, that the b quark couples predominantly to the c quark. Hence it came as a surprise when in 1983, the MAC and MARK II Collaborations at PEP, were able to infer that the lifetime was around 1 ps [4, 5]. Thus, the new measurement provided for the first time a value of the CKM matrix element $|V_{cb}|$ which is the transition strength between b and c quarks. Since then many lifetime measurements were performed with increasing precision due to more and more sophisticated detecting and reconstruction techniques. Very precise measurements were possible at LEP since 1993 after the development of high resolution silicon vertex detectors. This improvement made it also possible to confirm the theoretical prediction that the lifetimes of the various b-hadron species are different. From a naive point of view, where the other quark constituents play no role in the decay process *Spectator Model*, the lifetimes are expected to be equal. However, there exist processes which contribute to a hadronic width of the b-hadrons which are different for the various B species. A more detailed introduction to the theory of B decays and their description in the framework of Monte Carlo generators can be found in chapter one together with a brief description of the ideas and principles of the Standard Model.

At the LEP experiments various lifetime measurements have been made. Most of them were based on an exclusive reconstruction of certain decay chains which could be reconstructed with small backgrounds and which have a branching ratio which is high enough to have reasonably high statistics (for an example see [6]). Other analyses use the slow pion from the subsequent D^* decay to enhance B mesons of a certain type (for an example see [7]). In common to all these analyses is the statistical limitation imposed by ignoring most of the hadronic decay modes and thus using only a few percent of the available statistics. Over the years, however, improved detector knowledge and enhanced quality of simulation by using more and more precision measurements as inputs or constraints, has lead to the development of techniques which allow the inclusive treatment of B physics problems. Inclusive means that regardless of specific decay channels an approach was made to measure quantities from B physics using shapes of more general variables, like the weighted charge of the secondary vertex. In the Karlsruhe working group a dedicated B physics package named BSAURUS (see [94]) has been developed, providing a whole bunch of inclusive tools and variables some of which are basic inputs to this analysis. The package makes extensive use of all parts of the DELPHI experiment especially of its unique particle identification capabilities, and of elaborate neural network techniques. Chapter two gives a brief description of the LEP machine, DELPHI detector and its components. The first part of chapter three is dedicated to Neural Networks, their usage and some of their advanced training algorithms.

The inclusive approach of the analysis presented in this work allows the recovery of the large statistics of hadronic B decays for an inclusive b-hadron lifetime measurement based on the data taken by DELPHI in the years 1994 and 1995 (LEP I phase). The result is a new, very precise contribution to other lifetime measurements. For the measurement of the lifetimes of the B^+ and B^0 mesons[1] and for the measurement of the average b-hadron lifetime a χ^2 minimization technique was used. For the measurement of the lifetime of B_s and Λ_b the situation was quite difficult due to the relatively low production rates of these two b-hadrons ($\simeq 10\%$) and therefore the lifetimes had to be extracted against a huge background. This prompted to the decision to use a maximum log-likelihood

---

[1] Note that the corresponding charge conjugate state is always implied throughout this work if not stated otherwise.

method, where the background had to be modeled properly. An overview of the parameter estimation techniques, from a mathematical point of view, is given in the third part of chapter three. Using the maximum log-likelihood method the problem of modeling probability density functions arise and therefore calculations of numerical integrals are necessary. The B meson decay time is reconstructed from the decay length and the momentum of the b-hadron. Both quantities are reconstructed in the detector with limited resolution. The fitting method, which is later applied for the extraction of the results, takes these finite resolution effects into account by convoluting the resolution functions with the theoretical expectations. Therefore a numerical integration over these resolution functions is necessary for each event in the two dimensional phase space (decay length over momentum). The standard approaches for numerical integration (e.g. trapezoid rule) are found to be too slow in two dimensions to be of practical use when large statistics are involved and the precision is also not the best available. The standard Monte Carlo integration algorithms can be improved by two methods: First, the set of points used for the integration can be improved. Second, the integration error can be reduced by using so-called *Quasi-Random Numbers* instead of a grid or pseudo random numbers, for the same amount of points used for the integration. A brief introduction to quasi-random numbers is given in the second part of chapter three. Another improvement is the reduction of the variance of the function to be integrated. Technically, this is done by a variable transformation which transforms the quasi-random variable in a way according to the magnitude of the resolution function so that these points *see* a flatten function. This method is called *Importance Sampling.* A detailed description of this method is given in the second part of chapter three.

The basic properties of the package BSAURUS are described in the fourth chapter. Essential to the analysis is an estimate of the proper decay length, which implies a secondary vertex reconstruction and an estimate of the B momentum, from which the proper decay length can be calculated. This is achieved inside BSAURUS by the application of Neural Networks starting at the track level e.g. TrackNet which is dedicated to distinguish between fragmentation tracks and B decay tracks. Also quite essential for the analysis is the capability to distinguish between the different b-hadron species B^+, B^0, B_s and b-baryons. A neural network with four output nodes (one for each hypothesis) was constructed for this purpose.

In chapter five a study to measure b-hadron lifetimes using a maximum log-likelihood method is described. To measure inclusively the lifetimes of B_s mesons and b-baryons, which have a small relative production fraction of about 10% using this kind of method is essential. Performing a maximum log-likelihood method however requires a huge amount of precise modeling. The definition of the resolution classes of momentum and decay length, the modeling of the acceptance and the background are described in this chapter. Unfortunately it turned out, that the modeling was not precise enough and would require to take into account various correlations to have a robust analysis method. This turned out to be unpracticable with regard to the limited statistics available.

In chapter six, the analysis for the lifetime measurement of B^+ and B^0 mesons, based on a χ^2 method is presented starting with a description of the event selection. The analysis is based on data collected on the Z^0 resonance in the LEP I phase during the years 1994 and 1995. The analyses are done for the two years separately and combined afterwards because in 1995 some components of the particle identification were only running partly and so some adjustments and tunings had to be done separately for the two years. For each analysis the data were divided into two samples using the B species enrichment network, from which one of them was about 70% pure in charged B mesons and the other about 70% pure in B^0 mesons.

The chapter also contains a description of the fitting routine, as well as the calibration of known differences between data and Monte Carlo. Finally, after discussion of some modifications, the final results for the lifetimes and their ratio are presented, together with a breakdown and description of the systematic effects.

Chapter seven contains the description of the mean average lifetime measurement which follows closely to the method described in chapter five. The main difference is that there is no enhancement of specific b-hadrons and so the statistical error is quite small. Also here the final result and a breakdown of the systematic effects is given. The interpretation of the result is presented in terms of an estimation of the CKM matrix element $|V_{cb}|$

Finally chapter eight contains the conclusions of this work stating also all results with their errors.

Chapter 1

The Theoretical Background

Modern elementary particle physics studies the fundamental constituents of matter and their interactions (forces). Experimental discoveries and theoretical developments lead to a mathematical model, the so-called *Standard Model (SM)*. With help of this Standard Model a vast majority of known phenomena can be described with only a small number of particles which can be divided into two classes. The first class consists of quarks and leptons which are fermions of spin $\frac{1}{2}$ that obey Fermi-Dirac statistics. There are 3 charged leptons, three neutral leptons ('neutrinos') and six known types ('flavors') of quarks. The particles of the second class are the so-called 'gauge bosons'. They have spin 1 (the graviton has spin 2), obey Bose-Einstein statistics and are responsible for transmitting the fundamental forces between the particles. There are four known fundamental forces in the universe: the strong force binds quarks to form hadrons and nucleons to form nuclei, the electromagnetic force which acts on all electrically charged particles, the weak force, which is responsible for processes like radioactive beta decay and the gravitational force which is not yet included in the SM. The first part of this chapter gives an overview of the basic concepts of the Standard Model.

For an experimental analysis it is vital that there exists a good event simulation. Therefore understanding the process $e^+e^- \rightarrow$ hadrons is essential. In the second part of this chapter a brief overview of the simulation chain from a theoretical point of view is given.

The last part of the chapter describes the decays of heavy hadrons consisting of a c or b quark and a light (u,d or s) anti-quark. They decay via weak interactions. Using a naive model for the description one expects all b-hadrons (hadrons where the heavy quark is a b quark) to have the same mean lifetime. A more detailed look, however, shows that there are small differences in the decay width. A brief introduction to these phenomena and theoretical predictions is given in Section 1.4.

1.1 Gauge Theories

For the theoretical description of phenomena in elementary particle physics so-called Gauge theories are used. They are based on a generalized Lagrange formalism known from classical mechanics. The coordinates q_i in the Lagrange function $L(q_i, \dot{q}_i, t)$ and their derivatives $\dot{q}_i = \frac{dq_i}{dt}$ are substituted in quantum field theories by continuous fields $\Phi_i(x_\mu)$ and their derivatives $\frac{\partial \Phi_i}{\partial x_\mu}$.

$$\begin{aligned} q_i &\longrightarrow \Phi_i(x_\mu) \\ \frac{dq_i}{dt} &\longrightarrow \frac{\partial \Phi_i}{\partial x_\mu} \\ L(q_i, \dot{q}_i, t) &\longrightarrow \mathcal{L}(\Phi_i, \frac{\partial \Phi_i}{\partial x_\mu}, x_\mu) \end{aligned} \tag{1.1}$$

With this Lagrange density defined one can form the action S as

$$S = \int\limits_{-\infty}^{\infty} dx_\mu \mathcal{L}(\Phi_i, \frac{\partial \Phi_i}{\partial x_\mu}, x_\mu). \tag{1.2}$$

The dynamics of the fields are then obtained by minimizing S

$$\delta S[\Phi] = 0 \tag{1.3}$$

which is the well known variational principle from classical mechanics. From the variation of the physical fields Φ_i one can derive the Euler-Lagrangian equations of motion

$$\frac{\partial}{\partial x_\mu}(\frac{\partial \mathcal{L}}{\partial \frac{\partial \Phi_i}{\partial x_\mu}}) - \frac{\partial \mathcal{L}}{\partial \Phi_i} = 0 \tag{1.4}$$

Now one constructs $\mathcal{L}$ in a way that is invariant under local symmetry transformations $U(x)$ of the fields $\Phi_i(x)$

$$\Phi_i(x) \longrightarrow U(x)\Phi_i(x) \tag{1.5}$$

The structure of this underlying symmetry and its transformation dictates the kind of interaction that the theory describes between the fields Φ_i. In nature transformations under the $SU(n)$ group (the n-dimensional, special ($\det U = 1$) and unitary ($U^\dagger U = 1$) matrices U) are particularly relevant. Such a transformation has $n^2 - 1$ parameters and can be represented[1] as

$$U\Phi_i = \exp(i \sum_{a=1}^{n^2-1} \theta^a(x) T_a)\Phi_i \tag{1.6}$$

---

[1] We will adapt the Einstein sum convention where two equal indices implies a summation.

Here the $\theta^a(x)$ are the above mentioned parameters and the elements T_a are the generators of the corresponding group. They are not simple numbers but matrices. In general the following relation is valid for the generators:

$$[T_a, T_b] = i f_{ab}{}^c T_c. \tag{1.7}$$

Here the $f_{ab}{}^c$ are the structure constants of the group. In the case that these structure constants are all zero, the generators commutate with each other and the group is termed Abelian. The demand for invariance of the Lagrange density $\mathcal{L}$ under local gauge transformation requires the introduction of so-called gauge fields $W_\mu^a(x)$. They are inserted into the Lagrange density by substituting all partial derivatives ∂_μ by so-called covariant derivatives $D_\mu = \partial_\mu - igW_\mu(x)$ where g is the coupling constant of the corresponding interaction and $W_\mu(x)$ is an $n \times n$ matrix which can be expanded in the following way:

$$W_\mu(x) = T_a W_\mu^a(x). \tag{1.8}$$

As one can see from the equation above there are equal numbers of gauge fields and generators. Using a local gauge transformation according to Equation 1.6 the gauge fields have to transform as well to keep the Lagrange density invariant. This transformation has to fulfill the following condition:

$$D'_\mu \Phi'_i = U(x) D_\mu \Phi_i \tag{1.9}$$

This condition leads to the transformation of the vector fields

$$W'_\mu = U W_\mu U^{-1} - \frac{i}{g}(\partial_\mu U) U^{-1}, \tag{1.10}$$

resulting in

$$W'^a_\mu = W^a_\mu + \frac{1}{g}\partial_\mu \theta^a + f_{bc}{}^a W^b_\mu \theta^c \tag{1.11}$$

for infinitesimal gauge transformations. The gauge fields introduced by this procedure represent physical particles, the so-called gauge bosons. In an Abelian theory the gauge bosons do not interact with each other.

1.2 The Standard Model

The Standard Model describes the interaction of particles and is based on gauge theories. The elementary particles are ordered in 3 generations, each of them containing 2 quarks and 2 leptons. The main difference between these generations are the masses of the particles. The Standard Model contains the theory of

electroweak interaction by Glashow[8], Weinberg [9] and Salam [10] and the theory of quantum chromo dynamics (QCD) which describes the strong interaction. An overview of the particles, their division into generations and their quantum numbers can be seen in table 1.1
In the following subsections the components of the Standard Model are briefly summarized.

Generation			Quantum numbers			
I.	II.	III.	I	I_3	Y	Q
$\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L$	$\begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L$	$\begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L$	1/2	+1/2	−1	0
			1/2	−1/2	−1	−1
$\begin{pmatrix}u\\ d'\end{pmatrix}_L$	$\begin{pmatrix}c\\ s'\end{pmatrix}_L$	$\begin{pmatrix}t\\ b'\end{pmatrix}_L$	1/2	+1/2	+1/3	+2/3
			1/2	−1/2	+1/3	−1/3
e_R	μ_R	τ_R	0	0	−2	−1
u_R	c_R	t_R	0	0	+4/3	+2/3
d'_R	s'_R	b'_R	0	0	−2/3	−1/3

Table 1.1: *The different fermions in the standard model. Quarks and leptons are separated into three generations. Doublets are marked with an L to denote that only the left-handed components are arranged in this way whereas the righthanded components are arranged as singlets. The prime of the lower quarks in the doublets denote the weak eigenstate (see Section 1.2.2). The different quantum numbers are the electromagnetic charge Q, the weak hypercharge Y, the weak isospin I and the third component of the weak isospin I_3.*

1.2.1 Quantum Electro Dynamics (QED)

Quantum Electro Dynamics (QED) is the gauge theory describing electromagnetic interactions of charged fermions via the exchange of a single gauge boson, the photon γ. Since the photon itself is electrically neutral it does not self-interact. The underlying gauge group of this theory is the Abelian U(1) group. The coupling constant is proportional to the fermions electric charge which is a conserved quantity in all QED processes. Processes of particle physics can be illustrated via so-called 'Feynman diagrams'. These diagrams are not only graphical illustrations but also a symbolic way to represent mathematical terms of perturbation theory in different orders. As an example the electromagnetic process $e^+e^- \to \mu^+\mu^-$ is illustrated in Figure 1.1 (a). QED calculations to higher orders in perturbation theory than that represented by Figure 1.1 (a) involves the inclusion of closed-loop processes such as those shown in Figures 1.1 (b) and (c). These diagrams, however, correspond to divergent integrals which make infinite contributions to any amplitude calculations attempted beyond leading order. In QED it is found that these infinite terms may be absorbed by a redefinition of the charges and masses to be equal to their physically measured values. The application of this technique, known as 'renormalization', then renders all amplitude and hence cross-section calculations to be finite. The requirement that a candidate theory of particle interactions be renormalizable is clearly a prerequisite if physical predictions beyond leading order in the perturbation expansion are to be possible. It was shown by 't Hooft in 1971 [25] that, in fact, all local gauge theories are renormalizable.

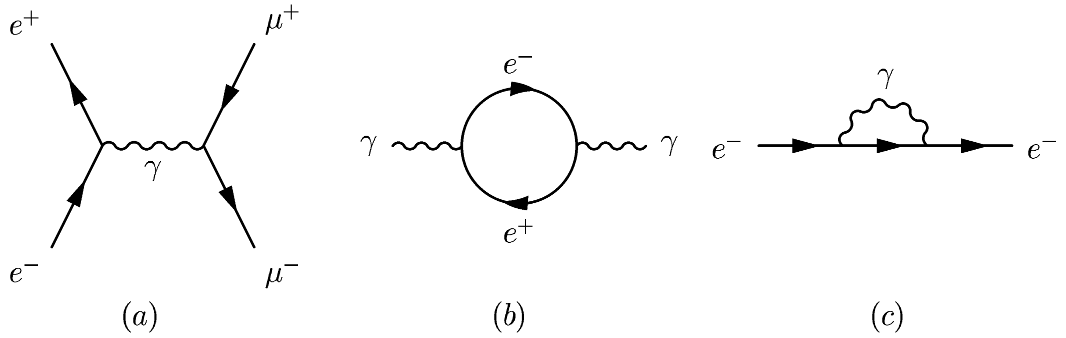

Figure 1.1: *(a) The lowest order (or Born level) QED-Feynman diagram for the process $e^+e^- \rightarrow \mu^+\mu^-$. Two examples of divergent diagrams in QED: (b) the creation and re-annihilation of an e^+e^- pair, and (c) the emission and reabsorption of a photon by an electron.*

1.2.2 Electroweak Interaction

The electromagnetic interaction and the weak interaction are unified in one theory, the electroweak interaction. The gauge group of this interaction is the $SU(2) \times U(1)$ group. The gauge bosons of the non-Abelian group SU(2) are W_1, W_2 and W_3. The physical charged vector bosons of the weak interaction W^+ and W^- are linear combinations of W_1 and W_2

$$W^{\pm} = 1/\sqrt{2} \cdot (W_1 \pm iW_2) \tag{1.12}$$

The gauge boson of the Abelian group U(1) is called B and couples to the hypercharge Y. The hypercharge Y is related to the electric charge Q and the third component of the weak isospin I_3 via the Gell-Mann-Nishijima formula:

$$Q = I_3 + \frac{Y}{2}. \tag{1.13}$$

The generators of the $SU(2)$ group are the Pauli spin matrices τ_a. The full $SU(2) \times U(1)$ invariant Lagrange density is then:

$$\begin{aligned}\mathcal{L} = \bar{\chi_L}\gamma^\mu(i\partial_\mu - g\frac{1}{2}\tau \cdot \mathbf{W}_\mu - g'\frac{Y}{2}B_\mu)\chi_L \\ +\bar{\Psi}_R\gamma^\mu(i\partial_\mu - g'\frac{Y}{2}B_\mu)\Psi_R \\ -\frac{1}{4}\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\end{aligned} \tag{1.14}$$

The two neutral physical gauge fields A_μ of the photon (γ)and Z_μ of the Z^0 can be represented by orthogonal combinations of the unphysical gauge fields $W_{\mu 3}$ and B_μ,

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \tag{1.15}$$

introducing a parameter θ_w known as the weak mixing angle or Weinberg angle. The trigonometric functions of this weak mixing angle are related to the coupling constants of the theory:

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \tag{1.16}$$

$$\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \tag{1.17}$$

Substituting $W_{\mu 3}$ and B_μ by A_μ and Z_μ in $\mathcal{L}$ (Eqn. 1.15) leads to:

$$\begin{aligned}\mathcal{L} = \bar{\chi_L}\gamma^\mu i\partial_\mu\chi_L + \bar{\Psi}_R\gamma^\mu i\partial_\mu\Psi_R \\ -\frac{g}{\sqrt{2}}(W_\mu^- \bar{\nu}_{eL}\gamma^\mu e_L + W_\mu^+ \bar{e}_L\gamma^\mu\nu_{eL}) \\ -\sqrt{g^2 + g'^2}Z_\mu[\frac{1}{2}\bar{\nu}_{eL}\gamma^\mu\nu_{eL} - \frac{1}{2}\bar{e}_L\gamma^\mu e_L \\ -\sin^2\theta_W \cdot (-\bar{e}_L\gamma^\mu e_L - \bar{e}_R\gamma^\mu e_R)] \\ -\frac{gg'}{\sqrt{g^2 + g'^2}}A_\mu(-\bar{e}_L\gamma^\mu e_L - \bar{e}_R\gamma^\mu e_R) \\ -\frac{1}{4}\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}.\end{aligned} \tag{1.18}$$

The hypercharges of the leptons are already inserted. The first two parts in the Lagrange density are the kinetic terms of the fermions and the last two parts are the kinetic terms of the gauge bosons and their self interaction because of the non-Abelian character of the gauge group. The parts in between describe the interactions of the fermions with the gauge bosons.

This electroweak theory does not contain any mass terms in its Lagrangian neither for the fermions nor for the gauge bosons. An introduction of a mass term would destroy the local gauge symmetry and therefore the renormalizability of the theory. To give the particles in the weak interaction theory a mass nevertheless the Ginzburg-Landau mechanism known from solid state physics is used. In this case this means that a new complex scalar field Φ, named Higgs field is used. By coupling the Higgs field Φ to the vector bosons the $W_{\pm}$ and Z^0 bosons become massive without destroying the symmetry. For the fermions Yukawa couplings to the Higgs field are introduced to give them masses. As a consequence of the Higgs mechanism a neutral physical boson H^0 with unknown mass is predicted[2].

Weak Quark Mixing

Within the electroweak theory, quark states can transform into their $SU(2)$ doublet partner through the emission of a $W^{\pm}$ boson which, in the absence of any mixing between the quark generations, renders the b-quark stable. The decay into the more massive top quark doublet partner is energetically impossible and all lighter quark flavors belong to $SU(2)$ doublets from different generations (see table 1.1). In the early 1960's though it became clear that the quark sector did not behave as simply as the lepton sector. The transition strengths for $e \rightarrow W\nu_e$ and $\mu \rightarrow W\nu_\mu$ are the same while the decay rate for $\pi^- \rightarrow \mu\nu_\mu$ and $K^- \rightarrow \mu\nu_\mu$ transitions differ by a factor of about twenty. A solution to this problem was then proposed by Cabibbo [12] who introduced a weak eigenstate for the down quark d' which is a linear combination of the mass eigenstates d and s in the following way:

$$d' = d\cos\theta_c + s\sin\theta_c \tag{1.19}$$

This way a mixing angle θ_c was introduced, the so-called Cabibbo angle which preserved the universality of the weak interaction by keeping one single coupling constant.

The problem of Cabibbos modification was, that flavor changing neutral currents (FCNC) were theoretically possible but were not experimentally observed. The solution of this problem was the modification of the s mass eigenstate into the weak eigenstate s' as well:

$$s' = s\cos\theta_c - d\sin\theta_c \tag{1.20}$$

To keep the coupling to the Z^0 universal, a new doublet for the s' state was introduced and therefore the existence of a new unknown quark state c was predicted, which is the so-called charm quark. This mixture of quarks ensures that flavor changing neutral currents (FCNC) can only occur in processes of higher order

---

[2] The current experimental lower limit of the Standard Model Higgs boson from the LEP2 collaborations is 114 GeV/c^2 [11]

perturbation theory. In agreement with experimental observations this guarantees that such FCNC processes are strongly suppressed. This mechanism is called the GIM (Glashow, Iliopoulios, Maiani [13]) mechanism. The charm quark was established experimentally by the discovery of the $c\bar{c}$ state in the mid 1970's [14].

In a later development Kobayashi and Maskawa proposed the extension of this to three generations to accommodate the possibility of CP violation introducing two quark states, the bottom b and the top quark t. According to this developement the wealth of experimental evidence for charged current transitions between generations, is theoretically accounted for by considering the weak quark eigenstates (d', s', b') as mixtures of the mass eigenstates (d, s, b)[3]. This can be expressed in terms of a unitary 3 × 3 mixing matrix due to Cabibbo, Kobayashi and Maskawa (CKM matrix) [15],

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.21}$$

The values of the CKM matrix elements are, like the Higgs mass, part of the free parameters of the standard model which cannot be predicted. Therefore the accurate measurement of the CKM matrix elements has great importance in constraining the uncertainty of predictions from electroweak theory. The current experimental values of the CKM matrix elements are [18]:

$$V_{CKM} = \begin{pmatrix} 0.9741 \text{ to } 0.9756 & 0.219 \text{ to } 0.226 & 0.0025 \text{ to } 0.0048 \\ 0.219 \text{ to } 0.226 & 0.9732 \text{ to } 0.9748 & 0.038 \text{ to } 0.044 \\ 0.004 \text{ to } 0.014 & 0.037 \text{ to } 0.044 & 0.9990 \text{ to } 0.9993 \end{pmatrix} \tag{1.22}$$

In general, the diagonal terms of the CKM matrix are close to unity whereas the off-diagonal terms are small but nonzero. The implication for decays of the b quark is therefore that transitions $b \to c$ and $b \to u$ can occur with amplitudes proportional to $|V_{cb}|$ and $|V_{ub}|$ respectively, i.e. at a rate slower than that for the process $t \to b$. The degree of mixing between the second and third generation is somewhat less than the degree of mixing between the first and second. This has the consequence that the lifetime of the b quark (and of any hadron containing a b quark) is longer than would perhaps be expected, and hence came as somewhat of a surprise when the first b-hadron lifetime measurements were made.

1.2.3 Quantum Chromo Dynamics (QCD)

The free Lagrange density for massless quarks (the mass of the quarks is again produced by Yukawa couplings to the Higgs boson) is

$$\mathcal{L} = \bar{\Psi}_c i \gamma_\mu \partial_\mu \Psi_c, \tag{1.23}$$

---

[3] Note that the (d, s, b) states are chosen to mix rather than (u, c, t) by convention only.

where the index c indicates that Ψ_C is a color triplet of the gauge group $SU(3)_c$

$$\Psi_c = \begin{pmatrix} \Psi_r \\ \Psi_g \\ \Psi_b \end{pmatrix} \tag{1.24}$$

introducing three color degrees of freedom (called red, green and blue). This color degree of freedom was introduced to resolve the so-called spin statistics problem with some baryonic states, such as the Ω, which contains three identical s quarks in a symmetric spin state. As a solution to the conservation of the Pauli principle each quark was assigned a different color resulting in a color singlet state. All experimentally observed states are color singletts.
The gauge principle for QCD requires the Lagrange density to be invariant under local $SU(3)$ transformation

$$\Psi_c(x) \longrightarrow \Psi_c(x)' = U\Psi_c(x) = e^{i\alpha^a(x)T_a}\Psi_c(x). \tag{1.25}$$

Here the generators T_a are 8 linear independent, traceless, hermitian 3×3 matrices called Gell-Mann matrices (see e.g. [16]) and $\alpha^a(x)$ the group parameters. After introducing the 8 Gauge fields G^a_μ like mentioned in Section 1.1 to ensure that the Lagrangian (Eqn. 1.23) is invariant an additional term $-\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a$ is added to the Lagrangian which represents the kinetic term of the gauge bosons (gluons). $G^a_{\mu\nu}$ is defined as $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + gf^a{}_{bc}G^b_\mu G^c_\nu$. All this leads to the following Lagrange density:

$$\mathcal{L}_{QCD} = \bar{\Psi}_c(i\gamma_\mu\partial_\mu - m)\Psi_c + g(\bar{\Psi}_c\gamma_\mu T_a\Psi_c)G^a_\mu - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \tag{1.26}$$

It describes the interaction of the colored quarks and the massless gluons which are also colored. Due to the structure of the field strength tensor $G^a_{\mu\nu}$ the term $-\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a$ does not only describe free gluons, but also the self interaction of gluons. The QCD Lagrangian (Eqn. 1.26) may be written symbolically in the following way:

$$\mathcal{L} = \bar{\Psi}\Psi + G^2 + g\bar{\Psi}\Psi G + gG^3 + g^2G^4 \tag{1.27}$$

The occurring terms describe in detail:

- the propagation of free quarks ($\bar{\Psi}\Psi$)
- free gluons (G^2)
- quark-gluon interaction ($g\bar{\Psi}\Psi G$)
- gluon self coupling (gG^3 and g^2G^4 (Fig. 1.2 (b) and (c)))

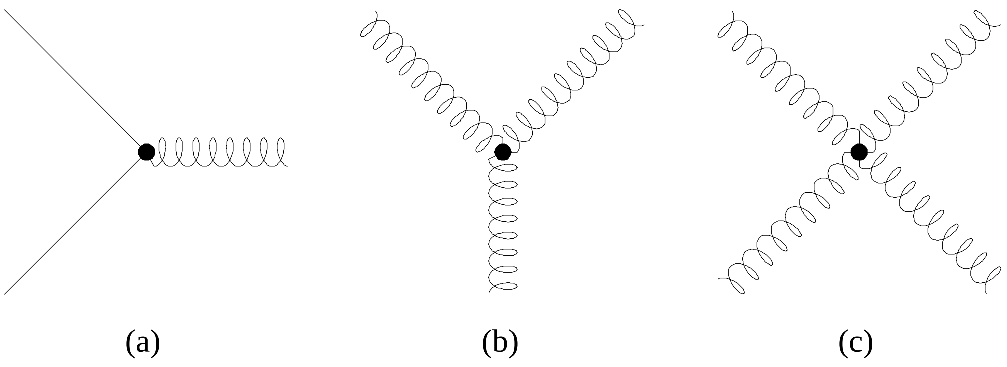

Figure 1.2: *The fundamental QCD vertices: (a) quark-gluon coupling, (b) triple-gluon-vertex, (c) four-gluon-vertex.*

1.2.4 Running Coupling Constants

As already briefly mentioned in Section 1.2.1 calculating higher order Feynman diagrams ('loops') like Figure 1.1 Figures (b) and (c) lead to divergent integrals, with an infinite contribution to any amplitude calculation beyond the Born level. As mentioned these divergencies vanish by redefining the masses and coupling constants to their physical values. As a consequence of those renormalization techniques a renormalization scale μ is introduced which can be interpreted as an energy scale. The coupling constant is not any more a constant, but depends on μ. $\alpha = \alpha(\mu)$ is therefore the coupling strength at the energy scale μ or at distance $1/\mu$. This means, that in order to have a theory with physical meaning the coupling constants have to be a function of the energy scale in which a problem is observed. This *running* of coupling constants has also been observed experimentally. The qualitative behavior of the coupling strength can be explained by simple physical reflection. In QED the electric charge of a particle is screened by its surrounding vacuum polarization (virtual $f\bar{f}$ pairs). The measured charge of this particle is increasing due to this 'screening effect'. In QCD not only fermion loops but, due to the non Abelian character of the strong interaction, also gluon loops contribute to the running of the strong coupling constant α_s. In this case the 'anti-screening effect' from the colored gluons overwhelms the 'screening effect' of the fermions. As a consequence the coupling constant of the strong interaction α_s is increasing with low momentum transfer. The running of the coupling constants leads to characteristic properties of QCD. For high momentum transfers α_s goes to zero. This behavior is called 'asymptotic freedom'. It permits the usage of perturbation theory for high momentum transfers. On the other hand α_s increases for low momentum transfers (big distances) which makes it impossible to calculate amplitudes using perturbation theory. And as a consequence of this quarks can only be observed in bound color singlet states.

This effect is called 'confinement'. In QCD the lower limit for momentum exchange q in a scattering process where perturbation theory is applicable is called Λ_{QCD}.

1.3 The Process $e^+e^- \to$ Hadrons

The e^+e^- annihilation is in contrast to $p\bar{p}$ and other hadronic annihilation reactions a process where the initial state is known with high accuracy. This makes e^+e^- collisions a good testbed for QCD. If the Z^0 created in e^+e^- collisions decays into a quark anti-quark pair the number of final state charged particles is 20 on average. In addition one observes in such multihadronic events several photons and neutrals. The particles are grouped in so-called jets. These jets approximately give the direction of the initial parton (quark or gluon).

The process which describes the evolution from the initial e^+e^- to the hadrons and leptons observable in the detector can be classified into four time ordered phases illustrated in fig. 1.3. The process starts with the electroweak production of a quark anti-quark pair via Z^0 or γ exchange (phase I). This is followed by a high energy strong interaction phase with parton cascade (phase II). This can be calculated using perturbation theory. In the third part low energy QCD and hadron formation takes place. This is not calculatable by perturbation theory and relies strongly on models (phase III). In the last phase the hadrons produced in phase III decay into final state stable[4] particles which can be observed in the detector (phase IV). We now discuss in more detail each of these four phases.

1.3.1 The Process $e^+e^- \to q\bar{q}$

The process $e^+e^- \to q\bar{q}$ in high energy collisions can be described by the Standard Model electroweak interaction in first order perturbation theory (known as the Born approximation or tree level) by the exchange of a photon or Z^0 boson. Neglecting the fermion masses the total cross section can be derived as [17]:

$$\sigma(e^+e^- \to q\bar{q}) = N_C \cdot \frac{4\pi}{3} \cdot \frac{\alpha^2}{s} \cdot \left[Q_e^2 Q_q^2 + (V_e^2 + A_e^2)(V_q^2 + A_q^2) \cdot |\chi|^2 + 2Q_e Q_q V_e V_q \cdot Re(\chi)\right], \tag{1.28}$$

where

$$\chi = \frac{1}{4\sin^2\theta_W \cdot \cos^2\theta_W} \cdot \frac{s}{s - m_Z^2 + i m_Z \Gamma_Z}. \tag{1.29}$$

---

[4] In this context stable means particles with a mean lifetime big enough to leave observable traces in the detector (e.g. $\pi^\pm$, $K^\pm$)

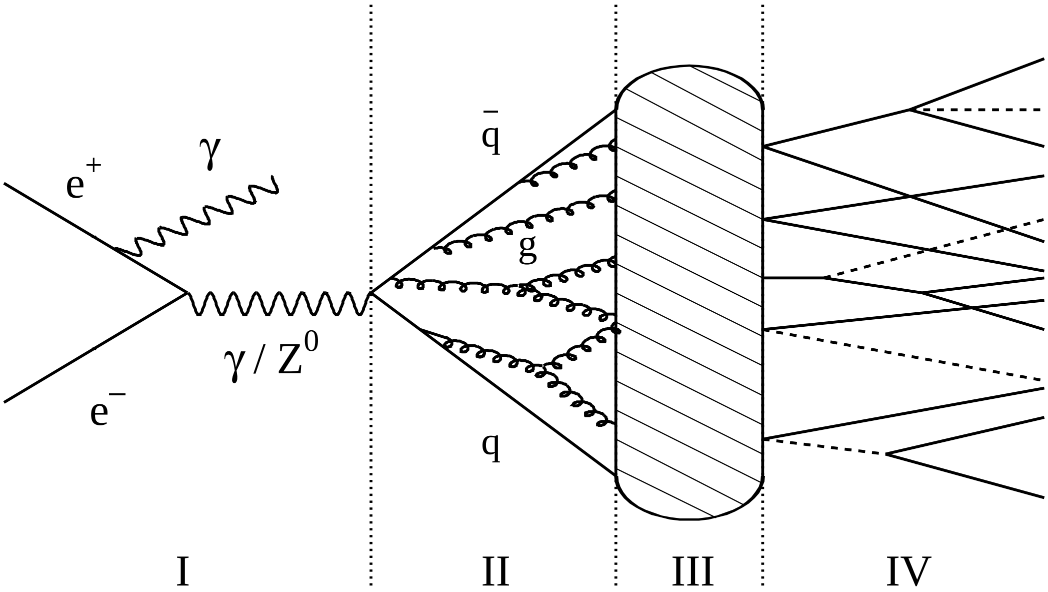

Figure 1.3: *The four phases of* $e^+e^- \rightarrow$ *hadrons: I) electroweak phase with initial state radiation, II) perturbative QCD phase based on gluon radiation and splitting as elementary processes, III) fragmentation and hadron formation phase (non perturbative QCD), IV) particle decays.*

$V_f = I_f^3 - 2Q_f \cdot \sin^2\theta_W$ and $A_f = I_f^3$ are the vector and axial-vector couplings of fermions to the Z^0, $\alpha = e^2/4\pi$ is the fine structure constant and $N_c = 3$ the number of colors. The first term describes the γ exchange, the second the Z^0 exchange and the third the γ/Z^0 interference. Formula 1.28 is modified by electroweak and strong corrections of higher order. These include virtual corrections (loop diagrams like Fig. 1.1 (b) and (c)), initial state radiation of photons and radiative corrections of photons and gluons in the final state. The dependence of the cross section on the center of mass energy and the relative contributions of the three terms can be seen in Figure 1.4. At a center of mass energy of 91 GeV, which is the mass of the Z^0 the Z^0 exchange is dominating. Therefore the relative production rates of various fermions are equal to the branching ratios of the Z^0 [18].

$$\begin{aligned} BR(Z^0 \rightarrow l^+l^-) &\approx 10.1\% \\ BR(Z^0 \rightarrow \nu\bar{\nu}) &\approx 20.0\% \\ BR(Z^0 \rightarrow q\bar{q}) &\approx 69.9\% \end{aligned} \tag{1.30}$$

The decay into neutrinos is not directly experimentally observable. Looking at quark production rates the up-type quarks (u and c) have a relative production rate of 17% whereas the down-type quarks (d, s and b) have a relative production rate of 22%.

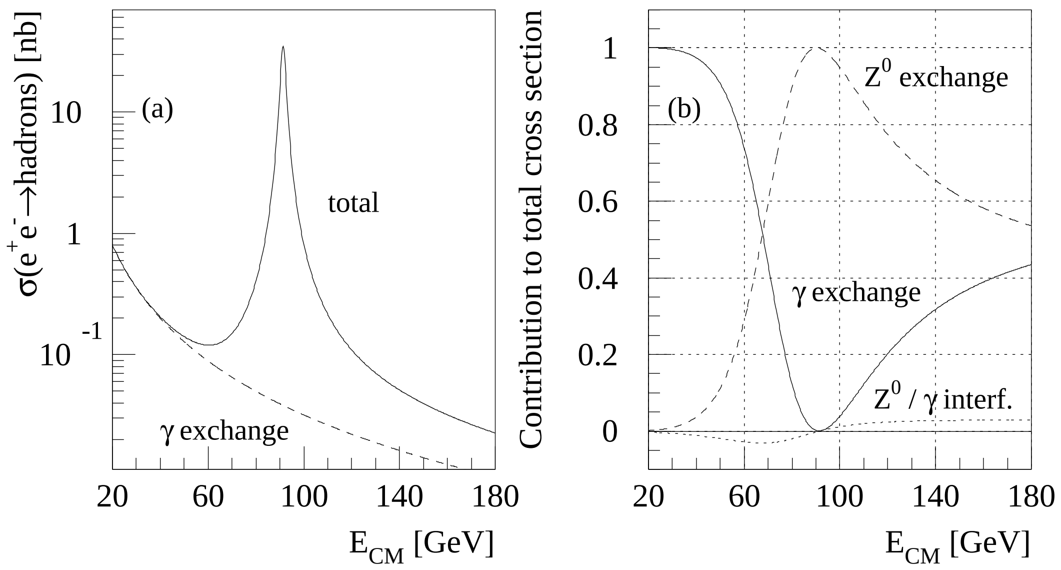

Figure 1.4: *Hadron production in* e^+e^- *annihilation. a) Total hadronic cross section as a function of center of mass energy. b) Relative contribution of* γ *and* Z^0 *exchange and of* γ/Z^0 *interference to the total cross section.*

1.3.2 The Perturbative Phase

Beginning with the quark anti-quark pair a configuration of color charged partons (quarks and gluons) is created in this phase by perturbative QCD processes. In Monte Carlo models two approaches are employed to describe this phase. The first method is the complete second order QCD matrix element (ME) calculation, and the second is the parton shower (PS) approach in leading log approximation.

QCD Matrix Elements

In this approach all QCD Feynman diagrams are calculated up to a certain order (until now complete calculations exist up to second order and tree level diagrams up to third order in α_s). In second order the maximum number of final state partons are four. This means that the final states $e^+e^- \to q\bar{q}$, $e^+e^- \to q\bar{q}g$, $e^+e^- \to q\bar{q}q\bar{q}$ and $e^+e^- \to q\bar{q}gg$ can be described.

Parton Shower Model

To perform exact multi hadron final states calculations, a huge amount of QCD Feynman diagrams have to be taken into account, which is practically impossible. An alternative to this approach is the so-called 'Leading Log Approximation'

(LLA). This method takes only the leading terms of each order into account. The multi parton configuration is reached by multiple application of the basic QCD processes $q \to qg$ (radiation of a gluon by a quark by bremsstrahlung), $g \to gg$ and $g \to q\bar{q}$ (splitting of a gluon into two gluons or a quark anti-quark pair). The probability that one of those processes $a \to bc$ takes place, is given by the Altarelli-Parisi[19] equations:

$$\frac{d\mathcal{P}_{a\to bc}}{dt} = \int dz \frac{\alpha_s(Q^2)}{2\pi}\mathcal{P}_{a\to bc}(z) \tag{1.31}$$

where $t = ln(Q^2/\Lambda^2_{QCD})$ denotes the evolution parameter and $\mathcal{P}_{a\to bc}(z)$ are the Altarelli-Parisi splitting functions [20]:

$$\begin{aligned} \mathcal{P}_{q\to qg}(z) = & \quad C_F\tfrac{1+z^2}{1-z}, \\ \mathcal{P}_{g\to gg}(z) = & \quad N_C\tfrac{(1-z(1-z))^2}{z(1-z)}, \\ \mathcal{P}_{g\to q\bar{q}}(z) = & \quad T_R(z^2+(1-z)^2), \qquad (1.32) \\ & \qquad (1.33) \end{aligned}$$

with $C_F = 4/3$, $N_C = 3$ and $T_R = N_F/2$. z or $(1-z)$ are the momentum fractions of the initial parton a transfered to the final state parton $b(c)$. The evolution of the shower continues up to a scale Q_0 which is a free parameter tuned to ≈ 1 GeV at LEP. Interference effects of higher order are taken into account by angular ordering of the radiated gluons. With parton showers one can get more than four partons in the final state which gives usually a better description of high energy data than the QCD matrix elements.

1.3.3 The Fragmentation Phase

The fragmentation (hadronisation) describes the creation of color neutral hadrons from color charged quarks and gluons. Because the perturbatively defined strong coupling constant becomes large at a Q^2 scale at about Λ^2_{QCD} this process is not accessible by perturbation theory (see Section 1.2.4). Commonly used fragmentation models are based on different physical assumptions. To describe real data a large number of free parameters in these models have to be fitted. Two models had great success in elementary particle physics in describing experimental data. One of those two models is the 'String Fragmentation' [21] which is used in the DELPHI Monte-Carlo generator. The model has several parameters which have to be extracted by a fit to the data. The parameters used for the DELPHI Monte Carlo for the years 1994 and 1995 are listed in Table 1.2. The other model, the 'Cluster Fragmentation' [22], which does not describe LEP data equally well, is only shortly described in this Section.

In general, both of the above mentioned methods are used. First, exact matrix element calculations are calculated to have a final state of four partons, followed by parton shower model.

String Fragmentation

In the model of string fragmentation a quark anti-quark pair move in opposite directions with velocity $v \approx c$. The self coupling of the color charged gluons in between those quarks lead to a linear field configuration or string with a typical length scale of 1 fm. The linear energy density within the color field is a constant of about $\kappa \simeq 1$ GeV/fm$\simeq 0.2$ GeV^2 which leads to a linearly growing potential between the quarks. If the field energy grows big enough that it is sufficient to create a new quark anti-quark pair, the string breaks up, leaving two separated $q\bar{q}$ pairs. They can build a meson or the described process happens again. This process is shown in Figure 1.5 schematically. In such a string fragmentation the created $q\bar{q}$ pairs get a fraction z respectively $(1-z)$ of the energy and longitudinal momentum. The probability density distribution of z is given by the longitudinal fragmentation function $f(z)$. Light quarks (u, d, s) are described in Delphi by the 'symmetric LUND function':

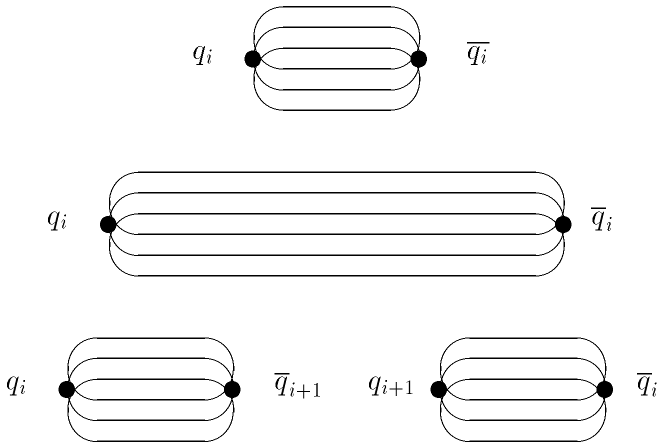

Figure 1.5: *Schematical picture of the string fragmentation. The primary $q\bar{q}$ pair moves apart. The energy of the field tube grows linearly with the distance until the string breaks up leaving two separate $q\bar{q}$ pairs. This process can happen repeatedly in those separate systems.*

$$f(z) = \frac{(1-z)^a}{z} \cdot e^{-bm_\perp^2/z}, \tag{1.34}$$

where $m_\perp^2 = m^2 + p_\perp^2$. a and b are free parameters which have to be fitted. Experimental data for heavy quarks (c and b) are described by the 'Peterson function' [23] in Delphi:

$$f(z) = \frac{1}{z(1 - 1/z - \epsilon_q/(1-z))^2}. \tag{1.35}$$

where ϵ_q denotes a free parameter, which is in principle the squared ratio of the masses of heavy to light quarks. Unfortunately the masses of the light quarks are not known well enough, so that ϵ_q has to be determined by comparison with experimental data. The parameters of the heavy quarks are then related: $\epsilon_c/\epsilon_b \approx m_b^2/m_c^2$. The fragmentation function for the different quarks is illustrated in Figure 1.6. There it can be seen clearly that the maximum for heavy quarks is located at high values of z. This behavior is called 'hard fragmentation'. This leads to the effect that the hadrons containing one of the primary heavy quarks carry most of the initial energy. The composition of quark flavors within the hadrons depends on the probability to create a certain quark flavor in the fragmentation process. This can be described by the quantum mechanical tunnel effect, because the massive quarks have to be created at a distance $2l$ where $l = m/\kappa$ for reasons of energy conservation. The tunnel probability is given by:

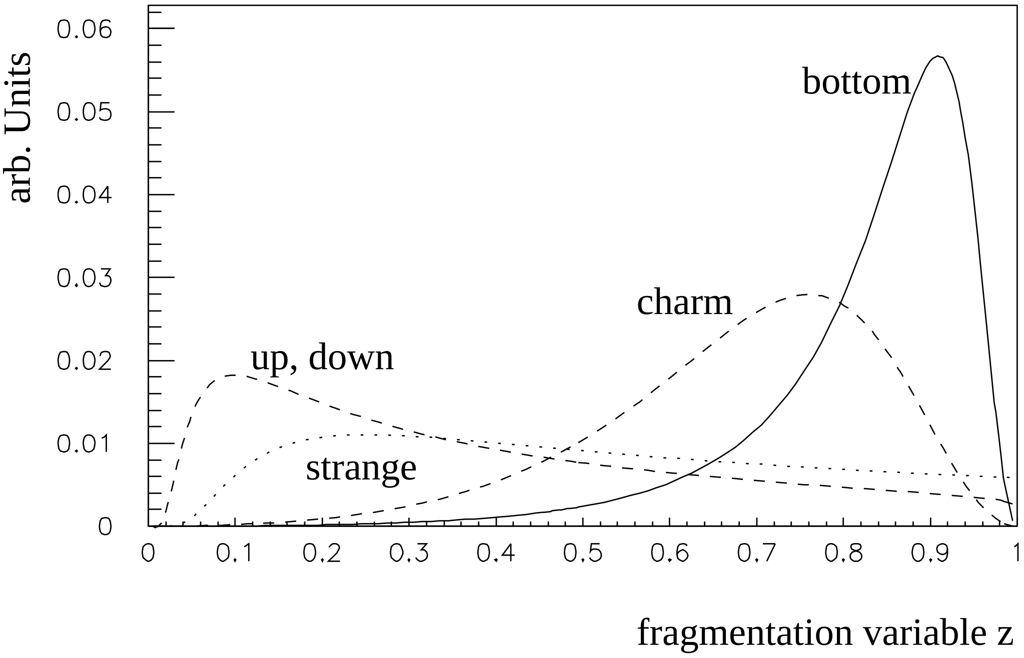

Figure 1.6: *The fragmentation function for different quark flavors.The heavy quarks have their maximum at high z values represented by their 'hard fragmentation'*

$$\mathcal{P} \propto \exp(-\pi \cdot \frac{m^2 + p_T^2}{\kappa}). \tag{1.36}$$

The p_T distribution is given in this model by a free parameter σ_{p_T}. From Equation 1.36 the relative rates are:

$$u : d : s : c : b \approx 1 : 1 : 0.3 : 10^{-11} : 10^{-100}. \tag{1.37}$$

Therefore the heavy quarks c and b are almost never produced in the fragmentation process, but frequently as primary $q\bar{q}$ pair in the electroweak γ/Z^0 exchange, in the perturbative QCD phase by gluon splitting or in weak decays $b \to c$. For the suppression of s quark production in fragmentation there exists a parameter γ_s which has to be fitted to the data. To simulate baryon production the creation of diquark pairs ($D\bar{D} = q_1q_2\bar{q}_1\bar{q}_2$) is inserted into the model. A baryon anti-baryon pair is then produced by combining the diquark with neighboring quarks. The production of diquark pairs is steered by several parameters. Their meanings are:

- the relative fraction of diquark production to $q\bar{q}$ pair production
- the suppression of diquarks containing $s\bar{s}$ pairs compared to diquarks containing only u and d quarks
- the suppression of diquarks with spin 1 compared to diquarks with spin 0

In addition to the production of baryons which are neighbors in the string, $q\bar{q}$ pairs in between the diquarks can produce mesons so that the baryon pair does not anymore succeed in the fragmentation chain. This process is called 'popcorn mechanism'.

Cluster Fragmentation

The principle of the cluster fragmentation is that after the parton shower which developed up to a certain energy scale Q_0 all gluons are forced to split into $q\bar{q}$ pairs. These quark anti-quark pairs are then collected into color neutral clusters. If the mass of such a cluster is bigger than a certain threshold the cluster decays into two new clusters. In the end the clusters decay, preserving color, spin and phase space constraints, into hadrons. The cluster fragmentation model has less parameters to adjust than the string fragmentation, but the data are less well described than using string fragmentation.
Another fragmentation model which will not be described here is the so-called model of 'independent fragmentation' [24], which is already ruled out by the data.

1.3.4 Particle Decays

The hadrons produced in fragmentation are either stable or they decay (direct or via cascades) into stable particles. In Monte Carlo generators measured branching

Parameter	**1994 Monte Carlo**	**1995 Monte Carlo**
LUND parameter a	0.354	0.417
LUND parameter b	0.523	0.850
σ_{p_T}	0.428 GeV	0.408 GeV
Λ_{QCD}	0.346 GeV	0.297 GeV
Q_0	2.25 GeV	1.56 GeV
ϵ_c	0.030476	0.03800
ϵ_b	0.002326	0.00284
γ_s	0.28	0.307
diquark rate	0.1	0.099
strange diquark suppresion	0.55	0.59283
spin 1 diquark suppression	0.07	0.07
popcorn rate	0.5	0.5
baryon suppression at end point of string	0.5	off
Z^0 mass	91.187 GeV/c^2	91.187 GeV/c^2
Z^0 width	2.489 GeV	2.489 GeV

Table 1.2: *The Table contains several key parameters of the Monte Carlo simulation which gave the best description of 1994 and 1995* DELPHI *data.*

ratios are used. For states, where some decay channels are not measured yet, predictions or observations from similar systems are used. This is the case in particular for b-hadrons. In the DELPHI simulation the decays of short lived particles (e.g. π^0) are implemented in the Monte Carlo generator. Particles which live long enough to interact with the detector are handed over to the detector simulation and decay afterwards.

1.4 Heavy Hadrons

Hadrons are named heavy hadrons if one of the quark constituents is a c or b quark[5]. At LEP they are mainly produced in the process $e^+e^- \to Z^0 \to c\bar{c}, b\bar{b}$ as explained in the previous section. Only a small fraction is produced by gluon splitting[6]. Hadrons containing a c quark can be produced either directly or via

---

[5] The top quark has a mass of $m_t = 175 \pm 6$ GeV/c^2 [26] and cannot be produced at LEP. In addition it does not produce hadrons, because the lifetime for the decay into a b quark is much shorter than typical hadronisation timescales.

[6] The rates to produce a $c\bar{c}$ or $b\bar{b}$ by gluon splitting have been measured at LEP:
$g_{c\bar{c}} = (2.27 \pm 0.28 \pm 0.41) \cdot 10^{-2}$(OPAL [27])
$g_{b\bar{b}} = (0.21 \pm 0.11 \pm 0.09) \cdot 10^{-2}$(DELPHI [28])
$g_{b\bar{b}} = (2.57 \pm 0.40 \pm 0.87) \cdot 10^{-3}$(ALEPH [29])

decays of b-hadrons. The center of mass energy of LEP is much higher than the threshold for $c\bar{c}$ or $b\bar{b}$ pair production and therefore a multitude of heavy mesons and baryons can be produced in either ground state or excited states. This allows hadron spectroscopy to be performed at LEP. Heavy hadrons, which have already decayed to their ground state, decay via weak interaction and can be schematically described by diagrams like Figure 1.7. As a consequence of the high mass a b-hadron has thousands of decay channels with small branching ratios which makes exclusive measurements difficult. As already mentioned in Section 1.3.3 heavy quarks have a hard fragmentation function, i.e. the primary c or b-hadron obtains a large fraction of the beam energy. b-hadrons have a mean energy of about $30 - 35$ GeV at LEP and a mean lifetime of about $\tau = 1.6$ ps. Therefore their mean flight distance is about $< l > = < \beta\gamma c\tau > \approx 3$ mm, where $\beta = v/c = pc/E$ and $\gamma = 1/\sqrt{1-\beta^2} = E/mc^2$. This can be exploited by high resolution micro-vertex detectors to tag $b\bar{b}$ events. In addition measuring the decay length allows to measure the mean lifetime of b-hadrons. This section gives an overview over theoretical tools to describe heavy hadrons. In addition theoretical predictions for the specific b-hadron lifetimes and a brief overview over the experimental status will be given.

1.4.1 Heavy Quark Effective Theory

For systems with one heavy quark where the limit $\Lambda_{QCD}/m_Q \rightarrow 0$ or $m_Q \rightarrow \infty$ is nearly reached, simplifications for their interactions can be made. Several studies in recent years, based on these assumptions [32], [33], [34] have led to a theory which is now commonly known as the 'Heavy Quark Effective Theory' (HQET). It provides a framework to make model independent predictions for heavy hadrons.

In a system with one extremely heavy particle, its dynamics can be kept constant compared to its light partners. Any new motion or energy level of the light partners will not change the situation of the heavy one. Following this picture a heavy quark in a bound state can be considered as a static source of color which is stable, whatever the light quarks are doing. Any change of the spin or radial state will be dominated by the light quarks. The equations of QCD in the neighborhood of such an isolated heavy quark are those of the light quarks and the gluonic degrees of freedom are dominated by the boundary condition of a static triplet source with the color-electric field at the origin. This leads to the expectation that the excitation spectra of heavy mesons and baryons will be nearly the same.

Since the spin also decouples in the heavy quark limit from the gluon field, all quarks look like scalar heavy quarks to their light partners. When flavor and spin of the heavy quark are irrelevant, the static heavy quark symmetry would lead to a $SU(2N_h)$, with N_h the number of heavy quarks. At the spectroscopic level this additional symmetry means that each spectral level of a heavy hadron will be degenerate.

The characteristics of HQET can be summarized in two pictures:

- A system with 'heavy flavor symmetry' is analogous to the chemistry of two isotopes with different numbers of neutrons. The electronic structure is equal, since they still have the same nuclear charge.

- The meaning of 'heavy spin symmetry' in this picture is an atom with degeneracy of the hyperfine levels. The resulting hyperfine splitting can be ignored, since the nuclear magnetic moments are small compared to the energy levels of the field interaction.

1.4.2 The Lifetimes of b-Hadrons

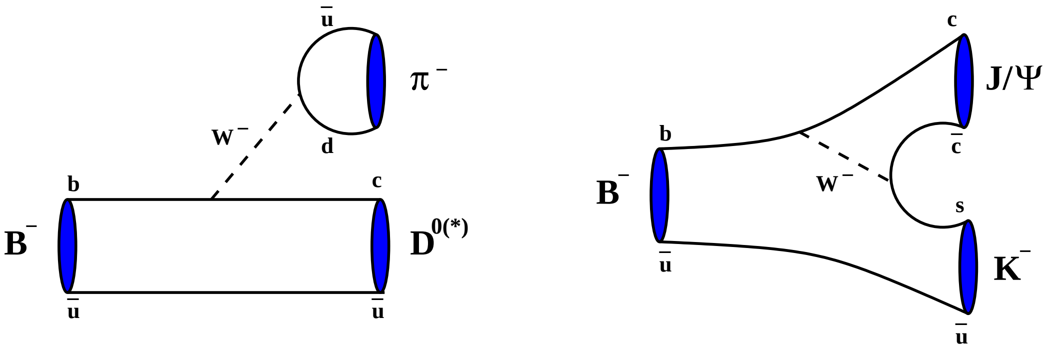

Figure 1.7: *Two examples for hadronic b decays in the 'Spectator Model'. The left diagram illustrates a weak decay of a B^- meson into a $D^{0(*)}$ and a π^-. On the right side the weak decay of a B^- into a J/Ψ and a K^- is visible.*

The b quarks produced in e^+e^- collisions build mesons and baryons in the fragmentation phase as described in Section 1.3.3. These hadrons are not stable and decay inside the detector. A lot of effort has been put into understanding the process of b-hadron decay and to predict the lifetime (see for example [36]).

Theoretical Overview

Calculating the mean lifetime of b-hadrons in perturbation theory one has to take into consideration two aspects. On one hand a b-hadron can decay semileptonically (see Figure 1.8) and on the other hand the W boson can decay into a quark anti-quark pair which can produce a meson (see Figure 1.7). Therefore the total decay width has a semileptonic and a hadronic fraction:

$$\Gamma_{tot} = \Gamma_{sl} + \Gamma_{had} \tag{1.38}$$

Introducing the semileptonic branching ratio BR_{sl}, the mean lifetime is:

$$\tau_B = \frac{BR_{sl}}{\Gamma_{sl}} \tag{1.39}$$

The simplest model to calculate the decay width is the so-called 'Spectator Model' in which the influence of the non decaying quarks in B^+, B^0, B_s and b-baryons is neglected, i.e. they are merely 'spectators'. The widths are calculated as if the b quark decays as a free particle. In this model all hadrons containing a b quark have the same mean lifetime. The matrix element of the semileptonic decay, in which the b quark decays either into a c or u quark can be written as the product of a leptonic and a hadronic current:

$$M_{sl} = -\frac{G_F}{\sqrt{2}} V_{qb} [\bar{q}\gamma_\mu(1-\gamma_5)b][\bar{l}\gamma_\mu(1-\gamma_5)\nu_l] \tag{1.40}$$

where G_F denotes the Fermi coupling constant of the weak interaction and V_{qb} the corresponding CKM matrix element. Since the third generation nearly decouples from the other two generations (small CKM matrix elements), the mean lifetime of b-hadrons is quite large ($\approx$ 1.6 ps). After squaring the matrix element and integrating over phase space (assuming all fermions massless) one gets the semileptonic decay width:

$$\Gamma_{sl}(b \to q) = \frac{G_F^2 m_b^5}{192\pi^3} |V_{qb}|^2 \tag{1.41}$$

which has the form of a modified muon decay width. In case of hadronic decay Equation 1.41 has to be modified by an additional scaling $|V_{ij}|$ (from the quarks of the W vertex) and a color factor $N_c = 3$. Naively, if we consider only the extra color factor associated with the hadronic decay channels, Equation 1.41 predicts:

$$\Gamma(W^* \to \bar{u}d : \bar{c}s : \bar{\nu}_e e : \bar{\nu}_\mu \mu : \bar{\nu}_\tau \tau) = 3 : 3 : 1 : 1 : 1 \tag{1.42}$$

where Γ is in units of $|V_{cb}|^2 G_F^2 m_b^5 / 192\pi^3$. The predicted semileptonic branching fraction is thus, $Br_{sl} = \frac{1}{9} = 11\%$, for each lepton flavor.

This rather naive picture of b-hadron decay can be extended by applying cor-

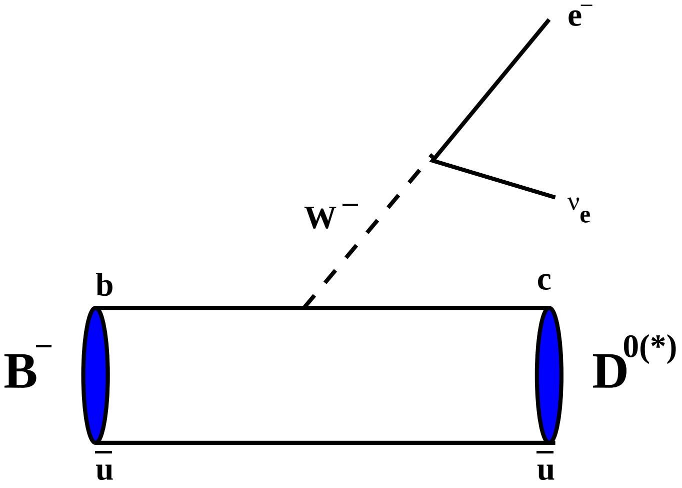

Figure 1.8: *In this Figure a semileptonic decay of a b meson is illustrated which decays into a electron neutrino pair and a* $D^{0(*)}$. *The branching ratio of* $B \rightarrow eX$, μX *is about* 11% *respectively, of* $B \rightarrow \tau X$ *it is smaller due to the high* τ *mass* ($\approx 2 - 3\%$).

rections to the spectator width of Equation 1.41. One of these corrections is the phase space suppression where a factor $F(\epsilon)$ is introduced to each partial decay width to account for non-negligible fermion masses compared to the b quark mass. Taking account of the direct decay quark from the b only, this factor evaluates to [36]:

$$F(\epsilon) = 1 - 8\epsilon^2 + \epsilon^6 - \epsilon^8 - 24\epsilon^4 \ln \epsilon \tag{1.43}$$

where $\epsilon = m_q/m_b$. Effective quark masses have been measured by CLEO and ARGUS, in the context of the ACM model [37], by fitting the spectrum of prompt leptons from b meson decays at the $\Upsilon(4S)$. The results are [39]:

$$\begin{aligned} m_b &= 4.95 \pm 0.04\ \mathrm{GeV/c^2} \quad (CLEO) \\ m_b &= 4.95 \pm 0.07\ \mathrm{GeV/c^2} \quad (ARGUS) \\ m_b - m_c &= 3.30 \pm 0.02\ \mathrm{GeV/c^2} \quad (ARGUS) \end{aligned} \tag{1.44}$$

In the case of $b \rightarrow u$ transitions, taking $m_u = 0.20$ GeV/c^2, the phase space correction is small and $F(\epsilon) = 0.99$. In the case of $b \rightarrow c$ transitions, however, and using the above ARGUS and CLEO mass results ($m_b = 4.95$ GeV/c^2, $m_c = 1.65$ GeV/c^2) we find that $F(\epsilon) = 0.45$. The suppression of $\bar{c}s$ and $\bar{\nu}_\tau \tau$ states is even higher. After applying the phase space correction, the total decay width, in the case of $b \rightarrow c$ transitions, is a factor of three smaller than that from the naive model, and BR_{sl} (for electrons and myons) increases to 16%.

To calculate the decay width even more accurately, QCD corrections have to be taken into account. There, one has to distinguish between soft gluon radiation like bremsstrahlung or vertex corrections and hard gluon exchange which rearranges the color indices of the participating quarks and therefore is color suppressed. In this case gluon lines go from one quark line to another. In Figure 1.9 both situations are illustrated in one picture. The correction factor for the soft gluon exchange can be expressed as [40]:

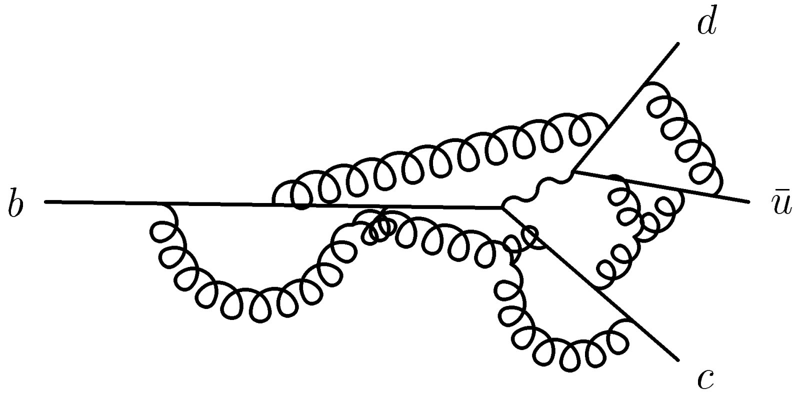

Figure 1.9: *This Figure illustrates the hadronic decay of a b quark with soft gluon radiation (e.g. vertex correction like gluon exchange between the d and the $\bar{u}$ quark) and hard gluon exchange (e.g. gluon exchange between the b and the d quark).*

$$F_{QCD} = 1 - (\frac{2\alpha_s}{3\pi}) \cdot g(\epsilon) \tag{1.45}$$

where we use the numerical approximation $g(\epsilon) = (\pi^2 - \frac{31}{4})(1-\epsilon)^2 + \frac{3}{2}$ [41]. This soft gluon correction has a similar form as the QED radiation corrections in the case of muon decay (substituting α_s for α). In the case of $b \to c$ transitions ($b \to u$ transitions can be neglected due to the CKM suppression[7]) this results in a soft gluon correction to the decay width of -10% for semileptonic decays and -4% for hadronic decays. The correction causes the predicted value of Br_{sl} to fall slightly to 15%.

In the case of hard gluon exchange, where the gluon lines rearrange the color indices such that the quark pair from the W decay no longer forms a color singlet, QCD corrections can be calculated [36]. These corrections act to modify the usual color factor of $N_c = 3$ and results in the hadronic decay modes being enhanced by approximately 13%. Since the semileptonic modes are unaffected by gluon exchange of this type, the semileptonic branching ratio is further

---

[7] This also implies that a measurement of τ_B is essentially a measurement of $|V_{cb}|$

reduced. Including all corrections up to this point we obtain a prediction from the corrected spectator model of $Br_{sl} = 14\%$.

The semileptonic branching ratio has been measured at the $\Upsilon(4S)$ (by ARGUS, CLEO, CUSB and Crystal Ball collaborations [42]) and at the Z^0 resonance at LEP [36]. The results are difficult to reconcile with the prediction of the corrected spectator model, which is in discrepancy with the data by (at worst) $\sim 20\%$.
The reason for the failure of the spectator model at this level, are largely associated with calculating the hadronic decay width. QCD interactions at the bottom mass scale are only marginally perturbative which leads to uncertainties associated with the calculation of hard and soft gluon effects as described above. A radiative gluon might have enough energy to split into a quark anti-quark pair. A fragmentation process might take place subsequently. Another problem is the fact that to describe the decay process correctly the quark fields should be replaced by corresponding hadron fields. The hadronic currents are therefore supplied with form factors which depend on the inner structure of the hadrons. These form factors are provided by effective potential models.
Most important for the discrepancy of predicted and measured semileptonic branching ratios however, is the fundamental assumption of the spectator model, that the spectator quark plays no role in the decay process. This can be a rather poor approximation when considering non-leptonic decay channels. Figure 1.10 shows some of the non-spectator diagrams that also contribute to b-hadron decay. These non-spectator effects result in different lifetimes for the different b-hadrons. There is another very important effect which contributes to a lifetime difference. If the final state contains two identical quarks, the Pauli principle holds and interference effects have to be taken into account. This is the so-called 'Pauli interference' (illustrated in Figure 1.11) which provides the largest contribution to the lifetime difference of the various b-hadrons. Theoretically, these effects are handled by an effective theory using an effective Lagrangian where the

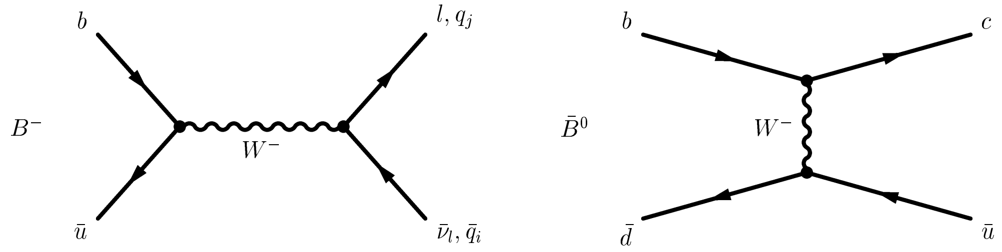

Figure 1.10: *Non-spectator diagrams for b-hadron decays. The left diagram shows the annihilation of a B^- meson. The right plot illustrates the direct W exchange in the decay of a $\bar{B}^0$ meson.*

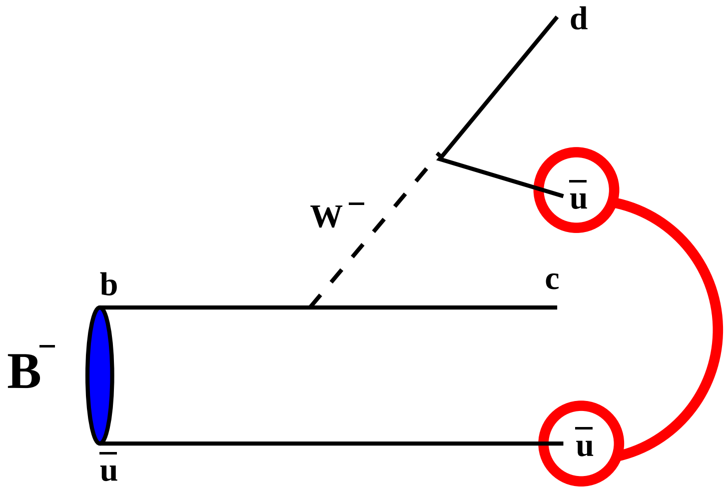

Figure 1.11: *In this Figure the Pauli interference is illustrated which occurs in the decay of charged b mesons. There are two identical particles in the final state which interfere and therefore contribute to a lifetime difference between charged and neutral b mesons*

interaction terms are simplified to local point interactions parameterized in different orders by so-called 'Wilson coefficients'. This is achieved by expanding the operators and currents in the Lagrangian in orders of $1/m_b$, the so-called 'Operator Product Expansion' (OPE) [36]. Calculating the decay width and expanding it in different orders of $1/m_b$ one recovers the above mentioned non-spectator effects.
The lifetime ratio of B^+ and B^0 is given theoretically by the equation [36]:

$$\frac{\tau_{B^+}}{\tau_{B^0}} \simeq 1 + 0.05 \cdot \frac{f_B^2}{(200\ \mathrm{MeV})^2} \tag{1.46}$$

where f_B denotes the theoretical decay constant which is of the order of 200 MeV. This is a result of HQET briefly described in the previous Section. One also expects

$$\tau_{B^0} \simeq \tau_{B_s} \tag{1.47}$$

No detailed analysis has been performed yet on τ_{Λ_b}; simple estimates lead to the expectation

$$\frac{\tau_{\Lambda_b}}{\tau_{B^0}} \sim 0.9 \tag{1.48}$$

Due to the fact that lifetime differences scale with orders of $1/m_Q$ b-hadron lifetimes do not differ as much as charmed hadrons because the mass is smaller.

The expectation for the hierarchy of lifetimes is

$$\tau_{\Lambda_b} < \tau_{B^0} \simeq \tau_{B_s} < \tau_{B^\pm} \tag{1.49}$$

There is a growing consensus between models that a difference in lifetime of order 5% should exist between the B^+ and B^0 meson whereas the difference between the B^0 and the B_s meson is predicted to be of the order of 1% [38]. The lifetime of b-baryons is predicted to differ from the lifetime of B^0 by about 10% [36].

The Experimental Status

In the last decade many experiments have provided measurements of the different b-hadron lifetimes and their ratios. The methods used delivered more and more precise measurements due to better analysis techniques and the use of silicon micro-vertex detectors (see Section 2.2.1). All these measurements, especially the newer ones, confirmed the hierarchy of lifetimes from the previous section. Figure 1.12 illustrates the current situation of the comparison of experiment and theory.

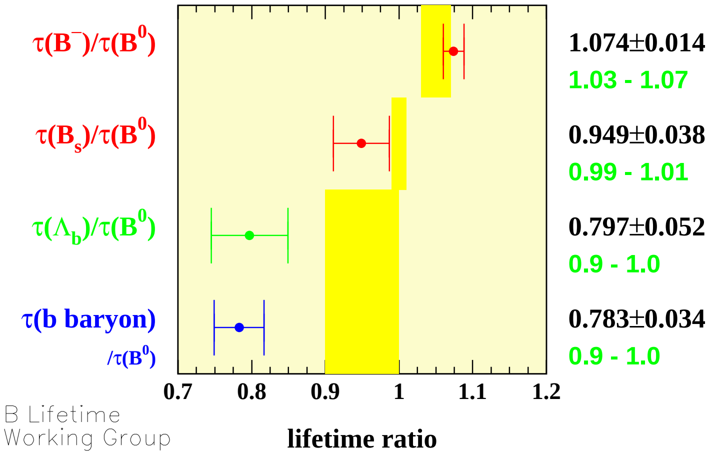

Figure 1.12: *Comparison of experimental results (world average) and theoretical predictions for the different b lifetimes. For b baryons the situation is unclear, a lot of contributions have not been calculated by theorists, yet.*

Chapter 2

The Experiment

The experimental environment of the LEP ring for the e^+e^- annihilations is briefly introduced. To understand the track reconstruction, the track chambers of the Delphi detector are described. Next the calorimeters and the particle identification devices are briefly described. At the end of this chapter the online and offline system including the data processing are explained.

2.1 The LEP Collider

The e^+e^- collisions took place at the Large Electron Positron storage ring (LEP) of the European Laboratory for Particle physics (CERN[1]) in Geneva, Switzerland. It was built from September 1983 until July 1989. Situated between the Jura mountains and Geneva airport, the main ring (see Figure 2.1) consisted of four straight sections, the interaction zones, and four arcs giving a total 'circumference' of 26.7 km constructed in an average depth of approximately 100 m. The machine parameters are listed in Table 2.1. At the four interaction zones the detectors ALEPH[2], Delphi [3], L3[4] and OPAL[5] were installed.

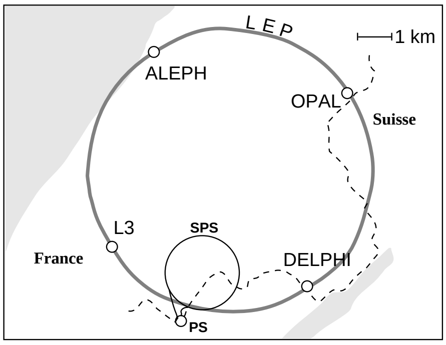

Figure 2.1: *Geographical location of the LEP collider.*

---

[1] 'Conseil Européen pour la Recherche Nucléaire'

[2] Apparatus for LEP PHysics

[3] DEtector with Lepton Photon and Hadron Identification

[4] Named after: 3rd Letter of intent

[5] Omni Purpose Apparatus for LEP

From 1989 until October 1995 LEP was running at a beam energy of about 45.6 GeV (LEP I phase) which in terms of center of mass energy corresponds to the mass of the Z^0 boson. By resonant production of this vector boson high precision tests of the standard model were possible which was the main goal of the LEP I phase.

Starting in October 1995 the center of mass energy was increased from 130 GeV up to 208 GeV by installing successively more superconducting cavities (LEP II phase). This has been done to enable the production of W pairs ($m_{W^\pm} \approx$ 80 GeV) and the search for new particles (e.g. Higgs bosons, supersymmetric particles). Further goals were the measurement of the energy dependence of Standard Model observables like the forward backward asymmetry A_{FB}.

In November 2000 all experimental LEP activities ended. The experiments were dismantled and the storage ring was removed. The LEP tunnel will be used for a new proton collider ring, the LHC[6] with a center of mass energy of about 14 TeV, which will be installed up to the year 2007. The main goals of the LHC are the discovery of Higgs boson and the search for new physics like supersymmetry. Two experiments are dedicated to this task, namely the CMS and the ATLAS experiments. A dedicated b factory (LHC-b) and a heavy ion experiment (ALICE) are also under construction. First data from proton collisions at the LHC are expected for 2008.

Circumference	26.66 km
Active length of RF structure	272 m
Injection energy	20 GeV
Maximum beam energy	55 GeV
Dipole field	0.0645 T
RF frequency	352 MHz
r.m.s. bunch length	15.7 mm
r.m.s. beam radii:	
σ_x	255 μm
σ_y	15.3 μm
Revolution time	90 μs
Bunch spacing	22 μs
Current per beam	3 mA
Number of particles per bunch	4.2×10^{11}
Nominal luminosity	1.6×10^{31} (cm^{-2}s^{-1})
Beam lifetime	6 h
Synchrotron radiation loss per turn	0.262 GeV

Table 2.1: *LEP (phase I) design machine parameters. From [43].*

---

[6] Large Hadron Collider

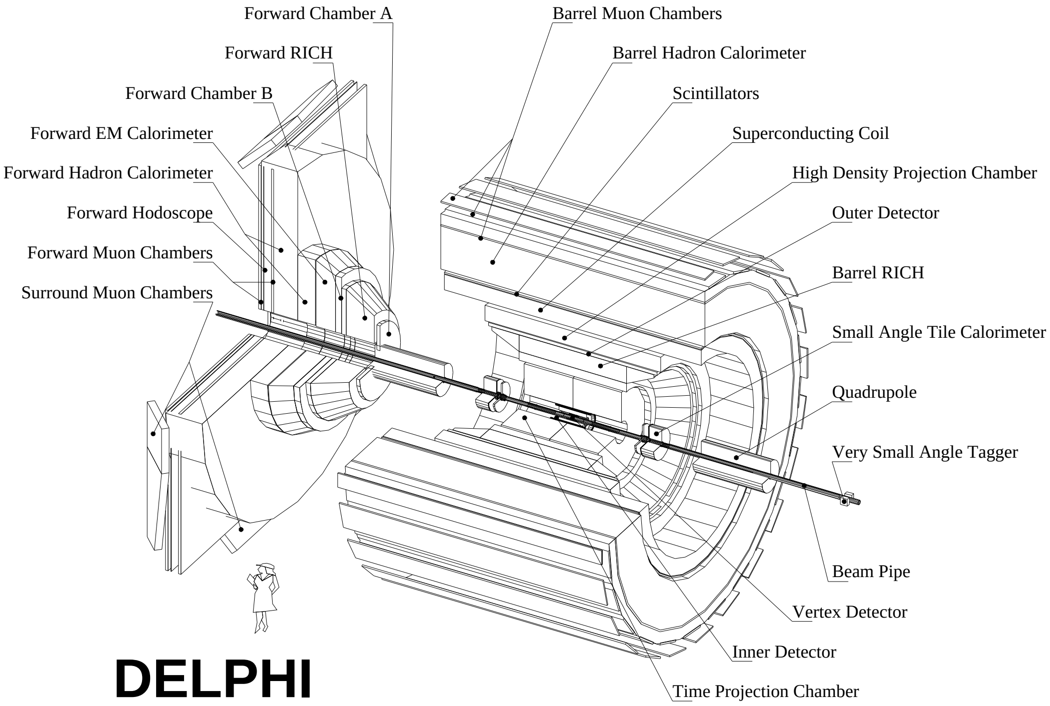

Figure 2.2: *Schematical view of the* Delphi *detector.*

2.2 The Delphi Detector

Delphi was designed as an omni purpose detector, the main goals of which were full covering of the solid angle of 4π, precise track reconstruction, good vertex reconstruction and an exact particle identification. The detector consisted of a cylindrical barrel part which was 10 m long and 9 m in diameter and an endcap on each side (see Figure 2.2). A coordinate system was defined such that the z axis started at the interaction point and pointed in the direction of flight of the electrons. The x axis pointed to the center of the storage ring and the y axis upwards. For simplification purposes cylindrical coordinates were used according to the geometry of the detector. θ was defined as the polar angle towards the z axis, ϕ was the azimuthal angle in the xy plane towards the x axis and $R = \sqrt{x^2 + y^2}$ was the distance to the coordinate origin in the xy plane. The barrel consisted of different subdetectors ordered in cylindrical layers around the interaction point. The individual tracking subdetectors were embedded in a magnetic field of 1.23 T from a superconducting magnetic solenoid, bending the tracks of charged particles in order to measure their momentum. A detailed description of the performance of the Delphi detector can be found in [44] and

[45].

In the following sections brief descriptions of the most important subdetectors of the barrel part are provided. Track and vertex reconstruction are more precise in the barrel by construction than in the endcaps.

2.2.1 Delphi Track Chambers

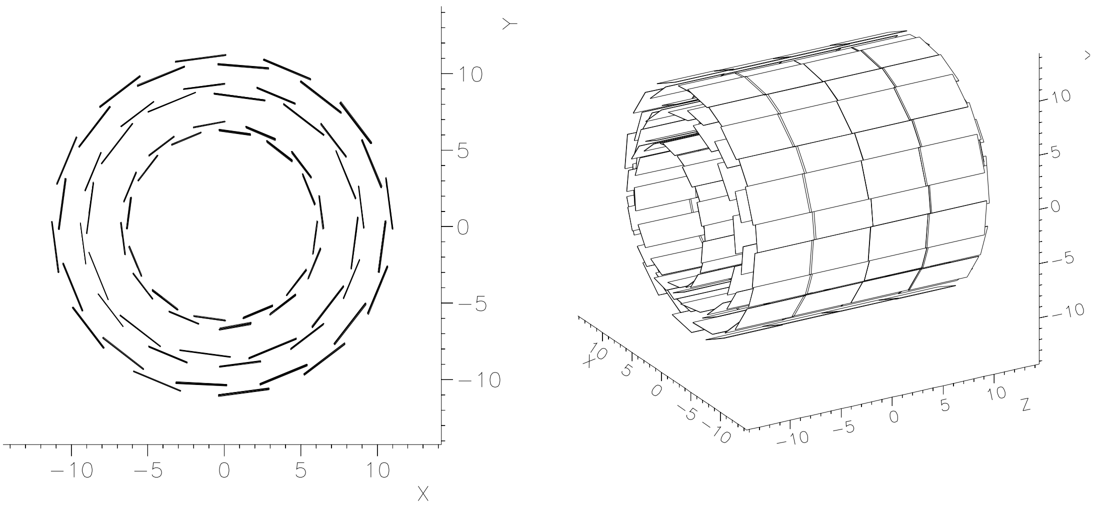

Figure 2.3: *Schematical View of the Micro Vertex Detector (widths in cm). Left: Projection into the plane orthogonal to the beam. Right: 3-dimensional view.*

The Micro Vertex Detector (VD)

The vertex detector was one of the most important detector components for b physics, because it was the detector closest to the interaction point and therefore it allowed the reconstruction of secondary decay vertices which is vital for tagging $b\bar{b}$ events and reconstructing them. The vertex detector consisted of three concentric layers of silicon strip detectors at radii of 6.3, 9.0 and 10.9 cm. Each layer consisted of 24 modules with an overlap in $R\phi$ of 10%. The inner layer had a length of 28 cm and covered the polar angle from 25° to 155°. The middle and outer layer had a length of 24 cm. Particles traveling at a polar angle between 44° and 136° were crossing each of the three layers of the vertex detector. Figure 2.3 shows a schematical view of the vertex detector. All three layers contained silicon strips parallel to the beam axis for precise measurement of the $R\phi$ coordinate of a track point. The inner and outer layer had additional strips perpendicular to the beam axis which allowed measurement of the z coordinate[7]. The resolution

---

[7] Here the configuration of the years 1994/1995 is described. Up until 1993 the VD only consisted of layers with $R\phi$ measurement.

of one single hit was 7.6 μm in $R\phi$. In z the resolution depended on the polar angle of the incoming particle because the number of hits increased with smaller angles. At $\theta = 90°$ the resolution was 9 μm.

The Inner Detector (ID)

The ID consisted of two parts. The inner part was a drift chamber with jet-chamber architecture pointing to the interaction point and the outer part consisted of five layers of cylindrical Multi-Wire-Proportional-Chambers (MWPC). The jet-chamber was divided into 24 sectors in ϕ and covered the region of $R = 12 - 23$ cm and $\theta = 23° - 157°$. Each MWPC layer consisted of 192 signal wires for the $R\phi$ measurement and concentric cathode strips for the z measurement. In 1995 the MWPC layers were replaced by Straw-Tube detectors which delivered no z information but covered an increased polar angular region of 15° up to 165°. The resolution of the ID was 50 μm in $R\phi$ and 1.5 mrad in ϕ. It allowed z measurement with a precision of 0.5-1 mm depending on θ. The ID was the most important event trigger in DELPHI with a fast readout within 3 μs.

The Time Projection Chamber (TPC)

The TPC was the biggest and most useful device for the measurement and reconstruction of charged particles. It started at $R = 28$ cm, covered about 3 m in z (with a sensitive length of ± 134 cm) and ended at $R = 120$ cm. Its active volume reached from $R = 35 - 111$ cm. As shown in Figure 2.4 the TPC was divided into two halfs (with six sectors in ϕ), which lied right and left of the interaction point. The electrostatic field of 187 V/cm in each half was directed to the end caps parallel to the beam axis and the magnetic field of the superconducting coil. The volume of the TPC was filled with an argon/methane (80%:20%) mixture. The ions produced by a passing particle drifted along the electric field parallel to the z axis and were read out at the end plates. Here 16 concentric pad rows and 192 anode wires were installed. In each of the 2×6 sectors were 1680 pads which formed the cathode rows. This allowed the measurement of 16 space points per track. A track passing through the gas volume left a tube of ionized gas along its way. The homogeneous electric field along the z axis lead to a drift of the ionization electrons towards the end-caps. The $R\phi$ position of a space point of a track was measured with the center of gravity of the charge clouds on cathode and anode with an accuracy of 250 μm. The z component was determined from the drift time compared with the trigger signal. The resolution in z was 880 μm. the separation of two tracks was possible if the tracks were further apart than 1 cm.

Besides the spatial information from the pads the sense wires provided up to 192 ionization measurements per track, which were used to calculate the specific ion-

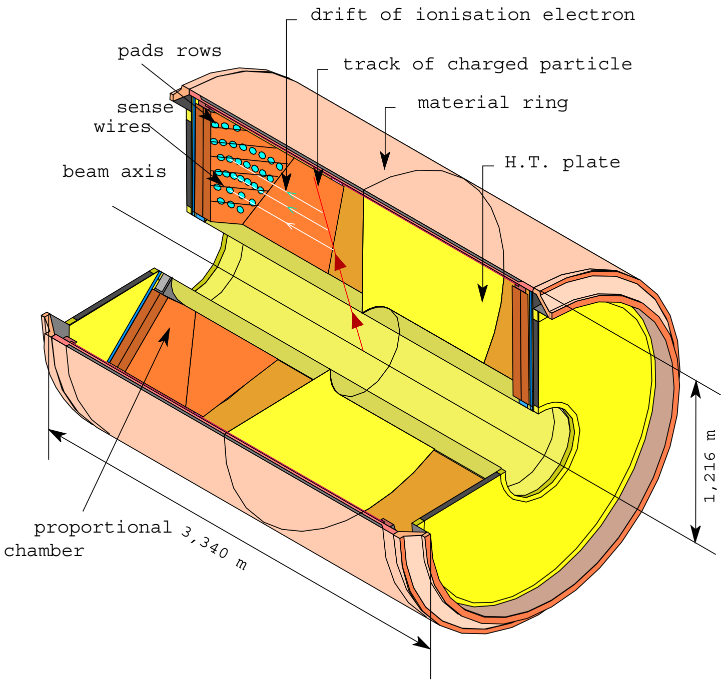

Figure 2.4: *The TPC geometry together with a track of a charged particle passing. The pad rows are indicated as circles the sense wires cross them as lines. A charge coming from the ionization of a track drifts towards the endcaps. The TPC was the most important tracking device of the* Delphi *detector.*

ization dE/dx of the tracks. This gave a contribution to the particle identification system for low momentum particles (< 1 GeV/c).

The Outer Detector (OD)

The Outer Detector lied between the barrel RICH and the electromagnetic calorimeter (HPC[8]). It was the tracking detector with the largest distance to the interaction point ($\approx$ 2 m). Thus the OD was able to improve the measurement of high energy tracks, which had a small curvature and completed the reconstruction of VD, ID and the TPC. In addition it was used for the matching of particle tracks to signals in the Cherenkov counters. The OD had a length of 460 cm and a polar angular coverage of 42° to 138°. The device consisted of 24 modules, which overlapped partly and allowed a complete coverage of the azimuthal region. Each module was made of five layers of drift tubes, which were parallel to the beam pipe between radii of 198 and 206 cm. The drift tubes allowed a good resolution in $R\phi$ of 110 μm and from the time differences of the

---

[8] High Density Projection Chamber

signals at the end of the chambers the z coordinate was able to be obtained with $\sigma_z = 3.5$ cm. The OD also contributed to the event trigger decision, as it had a fast readout of 3 μs.

The Forward Chambers A and B (FCA, FCB)

Besides the tracking devices in the barrel, in the forward and backward region of the detector other tracking devices, namely FCA and FCB were installed. The FCA ($|z| < 160$ cm, $11^\circ < \theta < 32^\circ$ and $148^\circ < \theta < 169^\circ$) was mounted on the end-caps of the TPC and was based on a straw tube technique. It reached a track resolution of 290 μm for the x and y direction. The FCB ($|z| < 275$ cm, $11^\circ < \theta < 36^\circ$ and $144^\circ < \theta < 169^\circ$) was a 12 layered multi-wire proportional chamber mounted on the FCA. It had a track resolution of 150 μm for the x and y direction. The chambers were used for the trigger decisions in the forward region and to expand the track reconstruction to smaller polar angles.

Tracking Performance

Results from all of the different tracking components were combined to derive more precise tracks. The alignment of the components and the disentangling of systematic effects (shifts, twists, etc.) are essential for good momentum resolution. Nevertheless, the main limitations came from the fact that a rather large volume of the space within the magnetic field was used by the Cherenkov detectors for particle identification (see Section 2.2.3). Using $Z^0 \rightarrow \mu^+\mu^-$ events the momentum resolution in the barrel part of the detector was determined to be:

$$\sigma(1/p) = 0.57 \times 10^{-3}(\mathrm{GeV}/c)^{-1}, \tag{2.1}$$

combining VD, ID, TPC and OD track elements [45]. The momentum resolution in the forward region ($20^\circ \leq \theta \leq 35^\circ$) was

$$\sigma(1/p) = 1.31 \times 10^{-3}(\mathrm{GeV}/c)^{-1} \tag{2.2}$$

using all available VD and FCB information.

2.2.2 Calorimetry

The High Density Projection Chamber

The High Density Projection Chamber (HPC) was the barrel electromagnetic calorimeter in DELPHI. It was the first large time projection gas calorimeter, which provides a full three-dimensional reconstruction of an electromagnetic

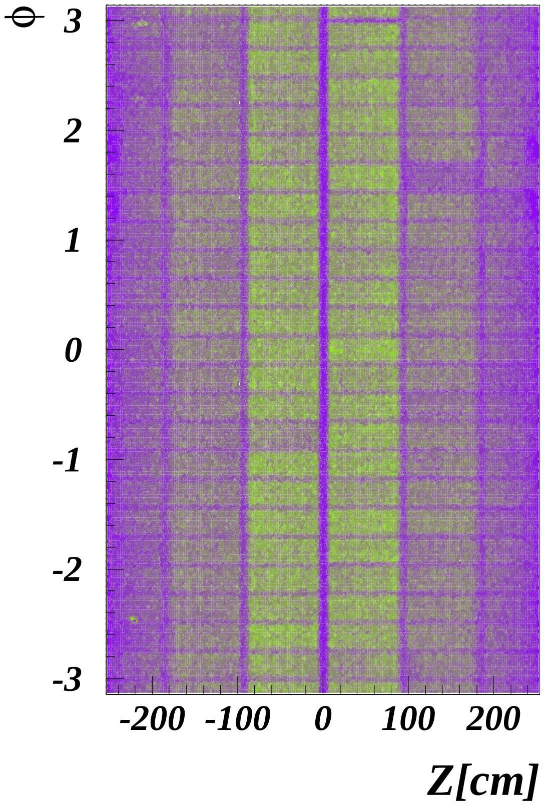

Figure 2.5: *The picture illustrates the occupation (the lighter grey the higher the occupation) of the HPC by neutral tracks in data back-traced to a radius of 217 cm which corresponds to the third layer in the HPC modules. The 144 individual modules are visible. One can also see modules which are nearly dead.*

shower. It covered the angular region $42° < \theta < 138°$. The HPC consisted of 144 modules arranged in 6 rings inside the cryostat of the magnet. Each ring consisted of 24 modules concentrically arranged around the z axis with an inner radius of 208 cm and an outer radius of 260 cm. In principle, each HPC module was a TPC, with 40 layers of a high dense material (lead) in the gas volume with a thickness of about 3 mm, where electromagnetic showers were initiated. The lead wires were not only used as converter material but also to build up the homogeneous electric drift field, parallel to the magnetic field. The converter thickness varied between 18 and 22 radiation length depending on the polar angle θ. The 39 gas gaps among the converter layers were filled with an argon/methane gas mixture (80%:20%). The produced cloud of electric charge formed an electromagnetic shower. It drifted with a velocity of $V_D = 5.5$ cm/μs in the above mentioned homogeneous electric field. The read-out of a single module was performed at the end of each module by a planar MWPC, which consisted of 39 sense wires and was segmented in 128 pads. Each pad was read out in 256 time buckets. This lead to the number of $144 \times 128 \times 256 = 4.7 \cdot 10^6$ ADC signals which were available per event. The granularity in $R\phi$ was 10 mm and 4 mm in z. The energy resolution of the HPC for photons had been found

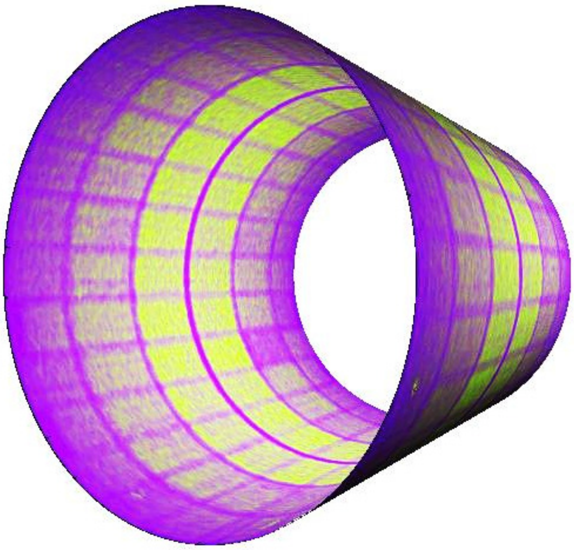

Figure 2.6: *Same plot as Figure 2.5 but for a scaled model of the third layer of the HPC modules.*

to be $\sigma_E/E = 32\%/\sqrt{E[\mathrm{GeV}]} \oplus 4.3\%$ (with $a \oplus b \equiv \sqrt{a^2+b^2}$) using neutral pions reconstructed from one photon converted before the TPC, reconstructed with high precision, and one photon reconstructed in the HPC [46]. The angular resolution in ϕ was given by the segmentation of the read-out ($\sigma_\phi = 3.1$ mrad). The z information has been evaluated from the drift time (leading to $\sigma_\theta = 1.0$ mrad). In each module additional plastic scintillators were installed, delivering a fast trigger signal. Figure 2.5 shows the occupation of the HPC from neutral tracks back-traced to the trigger layer at $R = 217$ cm which is the third layer. Figure 2.6 shows the same for a scaled model of the third layer.
The electromagnetic calorimeter in the forward direction (FEMC[9]) consisted of 9064 leadglass counters with approximately 20 radiation lengths, covering an angle of $1° \times 1°$ each in the polar angular region of $8° < \theta < 35°$ and $145° < \theta < 172°$. The energy resolution of the FEMC has been found to be $\sigma_E/E = 12\%/\sqrt{E[\mathrm{GeV}]} \oplus 11\%/E[\mathrm{GeV}] \oplus 3\%$

The Hadron Calorimeter

The return yoke of the Delphi superconducting solenoid was designed as an iron/gas hadron calorimeter (HAC). The angular acceptance of the instrument was $43° < \theta < 137°$ in the barrel part, and $11° < \theta < 50°$ respectively $130° < \theta < 169°$ in the two forward parts (HAF). The achieved granularity in ϕ direction was 3.75° respectively 3°(2.6°) in the θ direction of the barrel (forward). The active components of the HAC consisted of 20 layers (19 in the forward region) of plastic-made wire chambers of 2 cm depth interleaved by 5 cm of iron. The energy resolution obtained from multi-hadronic Z^0 decays using the momentum

---

[9] Forward ElectroMagnetic Calorimeter

information from the TPC was [46]

$$\frac{\sigma_E}{E} = \frac{112\%}{\sqrt{E[\mathrm{GeV}]}} \oplus 21\%. \tag{2.3}$$

The average depth of 1.5 m of the HAC together with the other detectors made DELPHI self-shielding so that even during runtime the detector remained accessible.

The Luminosity Monitoring Detectors

The luminosity measurement was covered by the SAT[10] and the VSAT[11] calorimeters installed in the extreme forward region. The luminosity measurement was done via low Q^2 Bhabha scatter events which emerge from the interaction point largely at small polar angles. The SAT consisting of lead and scintillating fibers was mounted at $z = \pm 2.35$ m. It was replaced in the winter break of 1993 by the STIC[12] a lead-scintillator sampling calorimeter read out with wavelength shifting fibres and photoelectrode tubes with a relative error of 0.2%, achieving a better resolution compared to the SAT. Another independent measurement was provided by the tungsten-sampling calorimeter VSAT, which was a W-Si calorimeter of 24 radiation length sensitive between the polar angles of 5 and 7 mrad, mounted at $z = \pm 7.5$ m.

2.2.3 Particle Identification

The Cherenkov Detectors

A property of DELPHI which distinguished it from the other three experiments was its dedicated framework for particle identification besides the energy loss measurement by the TPC.

The Ring Imaging Cherenkov (RICH) detectors were two detectors using the effect of Cherenkov light emitted by charged particles traversing a dielectric medium with a velocity larger than the speed of light in that medium. The emission angle θ_c depends on the mass m and momentum p of the particle via the relation

$$\cos\theta_c = \frac{\sqrt{1 + m^2/p^2}}{n}, \tag{2.4}$$

where n denotes the refractive index of the radiator medium. The number of the photons emitted is proportional to $\sin^2\theta_c$. This information (Cherenkov angle

---

[10] Small Angle Tagger

[11] Very Small Angle Tagger

[12] Small angle TIle Calorimeter

and number of photons) was used together with the momentum to evaluate the mass of charged particles. The main goal of the RICH detectors was to separate kaons and protons from the large pion background. In the momentum range where kaons and protons are below the Cherenkov threshold, they do not emit light while lighter particles do. This property had also been used in the so-called 'veto mode'.
The DELPHI RICH [47] contained two radiator systems of different refractive indices. A liquid radiator was used for particle identification in the momentum range from 0.7-4.0 GeV/c, and a gaseous radiator was used from 2.5-25.0 GeV/c. The full solid angle coverage was provided by two independent detectors (the forward and the barrel RICH). In Figure 2.7 the Cherenkov angle and the ionization loss $\frac{dE}{dx}$ is illustrated for different particle types. Perfluorocarbons were chosen as radiator media, both in the forward (liquid C_6F_{14}, gaseous C_4F_{10}) and in the barrel (liquid C_6F_{14}, gaseous C_5F_{12}). Photons in the range from 170-220 nm were focused onto photosensitive time projection chambers, 48 in number in the barrel and 24 in each arm of the forward RICH.

The Muon Chambers and Scintillator Counters

To identify muons which are able to traverse even big layers of material without interaction, two layers of drift chambers (one inside and one on the outer side of the hadron calorimeter) were installed. They provided a muon identification with an efficiency of about 95%. The Barrel, Forward and Surround[13] MUon (BMU, FMU and SMU) chambers cover the polar angular region between 15° and 165°. Resolution measurements on isolated tracks gave $\sigma_{R\phi} = 4$ mm. The z coordinate was evaluated from delay time measurements (with a digitization window of 2 ns) obtaining a resolution of $\sigma_z = 2.5$ mm.
The Time Of Flight counters (TOF) were installed on the outer surface of the solenoid. It consisted of a layer of 172 scintillation counters. The modules (19×2 cm^2 cross section and 3.5 m long) were read out at both ends by photomultipliers connected by light guides. In the forward part of the DELPHI detector a similar system was installed as well (HOF[14]). In the polar angular region from 15° to 165° the TOF system served as a cosmic muon trigger as well as a cosmic veto during beam crossings. Cosmic ray runs showed the time resolution to be 1.2 ns.

---

[13] The Surround Muon Chambers were installed in the winter break of 1994 to cover remaining gaps

[14] Forward HOdoscope

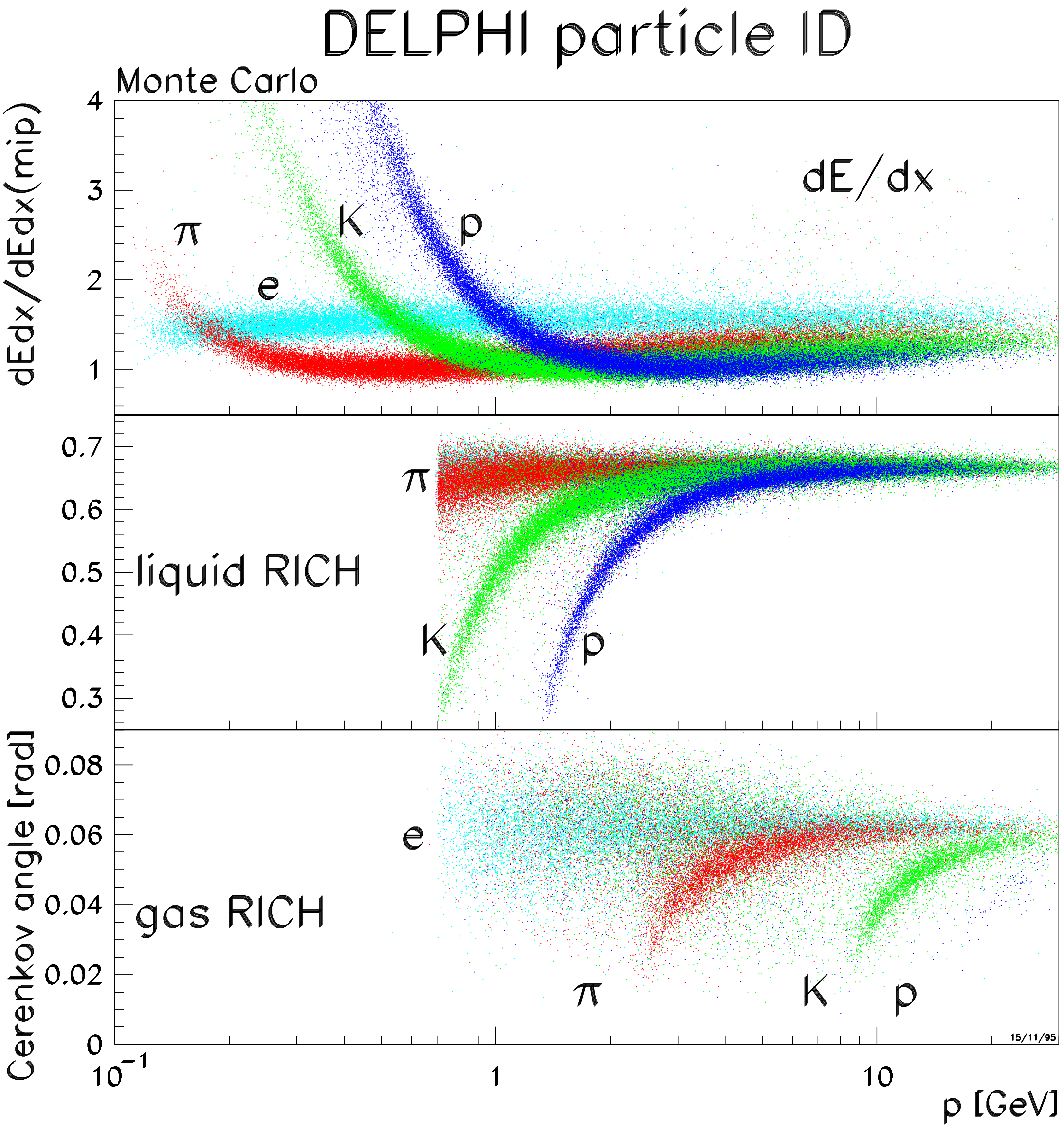

Figure 2.7: *dE/dx- and RICH information for particle identification taken from simulation. The energy loss dE/dx inside the TPC (top), and the Cherenkov angles in the liquid (middle) and gas (bottom) radiator for various particles.*

2.2.4 The Delphi Analysis Chain

The following section gives a short overview about the Delphi analysis chain. Starting with the slow control, trigger and data acquisition online system, then the DELANA, the DST and DELSIM offline system. For a detailed discussion see [48, 49, 50, 51, 52, 53].

The Online System

The Delphi online system had to manage several functions during runtime: analyze the events on an elementary level and supply a fast trigger decision; read out all detector components and write the data on storage media (disks, tapes); run the power supplies, gas and cooling systems of the detector and control and log all slowly varying detector parameters (e.g. temperatures, pressures, high voltages, drift velocities etc.).

The Delphi**slow control system** allowed a single operator to monitor and control the complete status (high and low voltages, gas supplies etc.) of the entire detector. The assignment comprised the display of the detector status, error messages and the continuous updating of the detector database for calibration and offline analysis. The 16 gas-filled subdetectors of Delphi were supplied by a standardized gas flow control system including automatic survey of relative mixtures, cleaning and drying of the media. This system was realized using VAX station computers shared with the Delphi data acquisition. The hardware link was realized by G64 computers for the subdetectors connected by an ethernet link.

The **trigger** was build in four levels (T1-T4), with the purpose to separate the events with physical interest from background events. The first two levels T1 and T2 were hardware triggers, which were synchronized with the Beam Cross Over (BCO). The time between two bunch crossovers was 22 μs in 4-bunch mode (11 μs in 8-bunch mode). Thus the first level trigger T1 had to be based on the detectors with a very fast readout (ID, OD, FCA, FCB) and the scintillators. In case of a positive first level trigger decision the slower T2 level analyzed in addition the informations of the slower drift chambers (e.g. TPC, HPC). T1 and T2 came to their decisions 3.5 μs respectively 39 μs after the bunch crossing. The next levels T3 and T4 were software based filters, using already a partial event reconstruction and particle identification. The trigger rates for T1 and T2 were 400 Hz and 5 Hz respectively. T3 and T4 reduced the background by an additional factor 1.5. The trigger system was installed with high redundancy to enable a stable running even with some components failing. The trigger efficiency of the system was extremely high and for multihadronic events close to 100% over the whole polar angle.

The **data acquisition system (DAS)** had the purpose to read out the digitized data from the detector and store them on magnetic tape. All events accepted by the second trigger level were stored on tape. For the raw data the data structure of the program package ZEBRA[15] [54] was used. The mean size of a multihadronic Z^0 event was around 150 kB. The events were identified by a 'Fill[16]' , 'Run[17]' and Event number.

The Offline System

This section gives an overview over the DELPHI offline analysis chain.

The main reconstruction program was the **Delphi ANAlysis program DELANA**. It contained one module for each subdetector which performs the necessary alignment and calibration of the raw data and a local track reconstruction if possible. A local pattern recognition was performed and resulted in e.g. track elements and energy depositions. These track elements were connected by a global search algorithm, starting at the TPC and then extrapolating to the inner and outer side of the detector. The found candidates were used to fit track parameters and to get rid of ambiguities. The tracks were then extrapolated through the whole detector and matching vertex detector hits were collected. With these hits the track fit was redone and a matching of tracks and neutral clusters was done. Finally, based on the fitted tracks of charged particles, a primary vertex was calculated.
The result of this procedure was a reconstructed event, stored in the DST[18] format. It contained informations used for physics analyses like momentum vectors, calorimeter information and others. The size of a multihadronic Z^0 event in the DST format is about 60 kB. The DST format contains enough information to calibrate the raw data without repeating the analysis process or to fit efficiencies in the simulation by smearing of resolutions.

An additional reduction of the data size was achieved by the production of **Short DST (SDST)** format out of DST. The aim of this step is the running of some standard reconstruction software which leads to savings in the CPU usage of the final analysis. The most important are:

- tagging of $Z^0 \to b\bar{b}$ events

---

[15] ZEBRA is a memory management program which provides dynamical data structures

[16] A 'Fill' denotes a LEP filling

[17] A 'Run' denotes a period of stable conditions for the detector

[18] Data Summary Tape

- identification of electrons, photons nd neutral pions
- identification of myons
- reconstruction of V^0 decays $K_s^0 \to \pi^+\pi^-$ and $\Lambda_0 \to p\pi$
- identification of charged hadrons ($\pi^\pm$, $K^\pm$, $p^\pm$)
- reconstruction of tracks using only VD hits
- reconstruction of interactions with the detector material

The removal of some individual detector information lead to a mean size of 20 kB in the SDST format of a multihadronic Z^0 decay. In a further step the data was reduced by an additional factor of three by storing SDST data into a **MINI-DST** format. The SDST format was replaced in 1996 by the **XSDST**, which contained some additional detector information.

The main task of the **detector simulation of Delphi (DELSIM**[19] [55] **)** was to produce simulated 'data' which mimic the real data from the detector as closely as possible. The output of the detector simulation were 'raw data' which were processed through the reconstruction program DELANA and the subsequent analysis programs in exactly the same way as for real data. DELSIM can be split into three different parts:

- The first part started with the physical generation of a primary physics process. Usually external Monte Carlo generators were used for this, e.g. JETSET [56], HERWIG [57], and ARIADNE [58] for $e^+e^- \to q\bar{q}$. The parameters in these packages were tuned from time to time with input from the latest available data.
- The particles produced were tracked through the detector until they hit material of the detector. This is done by stepping through the magnetic field including the possibility that these particles can undergo a secondary interaction. This required a detailed description of geometry and distribution of material for the different detector components. Besides energy loss and multi-scattering also photo-effect, compton-scattering, bremsstrahlung, positron annihilation, pair production, emission of δ-electrons, weak decays and hadronic interactions were taken into account.
- In the last step the detector response of each component was modeled. This part was built in a modular way, so that each detector component used a dedicated software package to take into account all specific properties of the underlying processes.

---

[19] DELphi SIMulation

The number of simulated Monte Carlo data events is typically larger than the number of real data events.

Chapter 3

Mathematical Tools

In high energy physics analyses often the problem arises to combine various variables in an appropriate way to separate a certain class of interesting events (signal) from a large amount of background. The simplest method to achieve this is to apply linear serial cuts on variables which are suitable for classification purposes. However, it turns out that such methods are inefficient and have a weak tagging power due to the neglect of correlations between the variables. In an inclusive analysis, like the one described in this work, more sophisticated methods have to be used. Common approaches are the Fisher discriminant analysis [59], the Likelihood Method [60] and artificial Neural Networks [61]. There are mainly three reasons why Neural Networks are prefered in this analysis: First a Neural Network uses every information about the correlations between the input variables and between the input quantities and the desired 'target'. These informations are coded in a set of free parameters, the weights w, which have to be adjusted by a training algorithm. These weights can be iteratively obtained by presenting examples of signal and background events to the Neural Network (training) together with the desired input. Second a Neural Network is more robust against statistical correlations between the input variables, which are not correlated with the desired target value, in contrast than e.g. the Likelihood Method. And third a Neural Network has the advantage that it is, once trained, very easy to handle and to propagate new events through the net to get the output. A brief mathematical introduction and overview of Neural Network techniques and their application forms the first part of this chapter.

The second part describes the numerical integration technique which is used for the Log-Likelihood method which will be described later. Based on a special set of numbers, so-called quasi random numbers a Monte Carlo integration is described making use of a sophisticated variable transformation to get a good precision while minimizing the necessary CPU time. The second part describes the statistical basics of parameter estimation used in this analysis.

3.1 Neural Networks

Neural Networks have seen an explosion of interest over the last few years and are being successfully applied for a vast range of problems in areas as diverse as medicine, finance, engineering geology and physics. Due to this occupation and research for algorithms to solve classification or prediction problems more and more sophisticated methods have been invented from which a small amount is presented here.

3.1.1 The Origin of Neural Networks

Neural Networks grew out of research in artificial intelligence. Attempts were made to learn of biological neural systems by modeling the low-level structure of the brain. A brain consists of about 10 billion neurons each of them connected to several thousands other neurons. A single neuron is a specialized cell with the ability to propagate an electrochemical signal. It has a branching input structure (the dendrites), a cell body, and a branching output structure (the axon). The axon of one cell connects to the dendrites of another via a synapse. When a neuron is activated it 'fires' an electrochemical signal along the axon. This signal crosses the synapses to other neurons which may fire in turn. A neuron fires only if the total signal received at the cell body from the dendrites exceeds a certain level (the firing threshold). Therefore changing the firing conditions means changing the 'strength' of synaptic connections and so learning consists of altering these strengths. Thus from a very large number of extremely simple processing units the

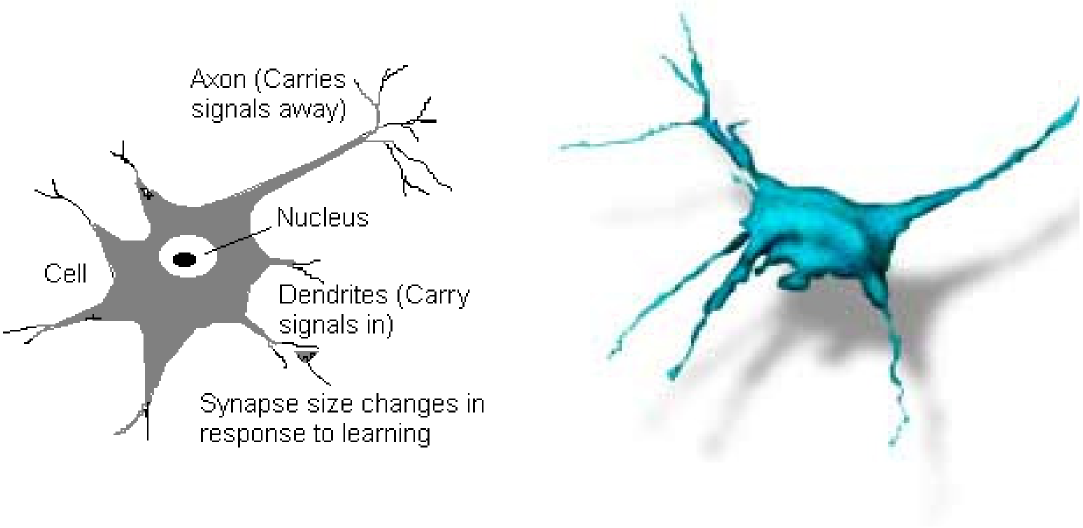

Figure 3.1: *The left plot is an illustration of the properties of a biological neuron. The right plot shows the three dimensional structure of a neuron.*

brain manages to perform extremely complex tasks. Artificial Neural Networks mimic those basic cells and achieve remarkable results using a model which is no more complex than this. Figure 3.1 illustrates the above mentioned properties of biological neurons.

3.1.2 The Basic Concept

To apply the model of biological neural systems to artificial neurons one constructs a mathematical model of such a neuron with exact defined and relatively simple properties. Each basic processing unit or neuron receives input x_i from neighbors or external sources and uses this to compute an output signal which is propagated to other units. Apart from this processing, a second task is the adjustment of the weights w_i which correspond to the connection strengths of the synapses. In addition, a neuron also has a single threshold w_0, called the bias which represents the threshold for the neuron to fire. The weighted sum of the inputs is formed and the threshold added to compose the activation of the neuron

$$y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 \tag{3.1}$$

which is activated if $y(\mathbf{x}) > 0$ (which corresponds to a step function as activation function). This model of a single neuron is illustrated in Figure 3.2.

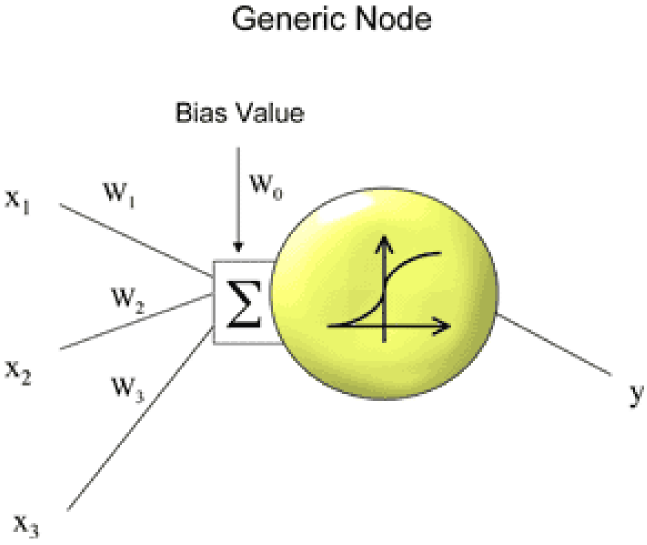

Figure 3.2: *The picture at left represents a generic node of an artificial Neural Network. The sigma represents the weighted sum of the input variables x_i. The S-shaped curve in the circle represents the second part of the processing, the passing through a nonlinear activation function.*

To generalize the linear discrimination function 3.1 a nonlinear activation function $\mathcal{F}(y)$ can be introduced, acting on the linear sum. Generally $\mathcal{F}(y)$ is chosen monotonic and in the simple case described above its a simple step function. Nonlinear activation functions are what give Neural Networks their nonlinear capabilities. One of the most common forms of activation function is the sigmoid which is a monotonically increasing function that asymptotically approaches some finite value as ±∞ is approached. The most common examples are the standard logistic function $\mathcal{F}(y) = 1/(1 + e^{-y})$ and hyperbolic tangent $\mathcal{F}(y) = tanh(y)$. Sigmoids that are symmetric about the origin are preferred because they are more likely to produce outputs (which are inputs to the next layer) that are on average close to zero[1], in contrast to the logistic function whose outputs are always positive and so must have a mean that is positive. The motivation to choose a logistic function however is based on the fact that it can be derived by Bayes' theorem, where the posterior probability of membership of class C_1 is given by

$$P(C_1|\mathbf{x}) = \frac{P(\mathbf{x}|C_1)P(C_1)}{P(\mathbf{x})} = \frac{P(\mathbf{x}|C_1)P(C_1)}{P(\mathbf{x}|C_1)P(C_1) + P(\mathbf{x}|C_2)P(C_2)} \quad . \tag{3.2}$$

If the class conditional probabilities $P(\mathbf{x}|C_i)$ are given by Gaussian distributions Equation 3.2 is equal to the logistic function which allows thus the output of the discriminant to be interpreted as posterior probabilities[2]. The logistic function can also have an additional parameter T to become:

$$\mathcal{F}(y) = \frac{1}{(1 + e^{-\frac{y}{T}})} \tag{3.3}$$

The parameter T is comparable, according to the Boltzmann distribution, with the (synthetic) temperature of the system. This 'effective temperature' is used in so-called Boltzmann machines which are similar to simulated annealing. The network starts with a very high value in T which is slowly decreased as the network iterates through the process of updating the nodes and weights (training). The hope is that the network escapes from local minima in the training and finds the global minimum. The activation function 3.3 is plotted for several temperatures in Figure 3.3.

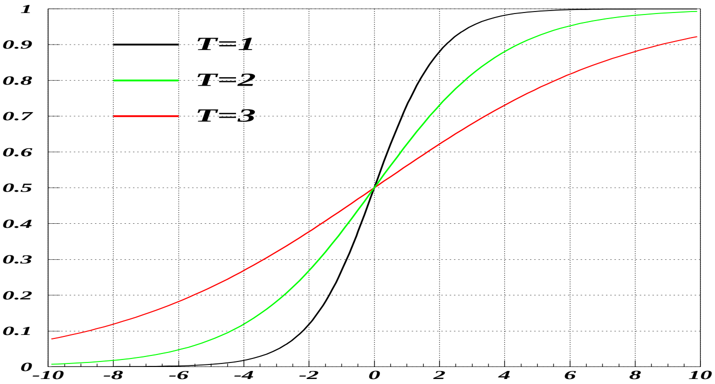

Figure 3.3: *This plot illustrates the nonlinear transfer function with an effective temperature according to the Boltzmann distribution for three different temperatures.*

---

[1] The reason for this is given in the following Sections.

[2] This argument holds only if the class-conditional probabilities are normally distributed with equal covariance matrices.

3.1.3 Architecture

In the previous Section the properties of the basic processing unit in an artificial Neural Network was discussed. This section focuses on the pattern of connections between the units and the propagation of data.As for this pattern of connections, the main distinction we can make is between:

- *Feed-forward* networks, where the data flow from input to output units is strictly feed-forward. The data processing can extend over multiple (layers of) units, but no feedback connections are present, that is, connections extending from outputs of units to inputs of units in the same or previous layers.
- *Recurrent* networks that do contain feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such that the network will evolve to a stable state in which these activations do not change anymore. In other applications, the change of the activation values of the output neurons are significant, such that the dynamical behavior constitutes the output of the network (see [62]).

Classical examples of feed-forward networks are the Perceptron and Adaline. Examples of recurrent networks have been presented by Anderson [63], Kohonen [64] and Hopfield [65] and will not be further discussed here because they do not have much relevance in high energy physics and have not been employed in this analysis.

A single layer feed-forward network (*single-layer* perceptron) consists of one or more output neurons o, each of which is connected with an adaptive weighting factor w_{io} to all of the inputs i. The input of one output neuron is the weighted sum of the inputs plus the bias term, which can be regarded as an additional input node with input value 1 and the threshold as weight. The output of the network is formed by the activation of the output neuron considered, which is some function of the input:

$$y_o = \mathcal{F}(\sum_{i=1}^{N} w_{io} x_i + \theta) \tag{3.4}$$

The activation function $\mathcal{F}$ can be linear, or nonlinear. Considering the threshold[3] function

$$\mathcal{F}(s) = \begin{cases} +1 & \text{if } s > 0 \\ -1 & \text{otherwise} \end{cases} \tag{3.5}$$

---

[3] Also the Heaviside or the sgn function is commonly used.

the output of the network is either +1 or -1, depending on the input and it can now be used for a classification task: it can decide whether an input pattern belongs to one of two classes. If the total input is positive, the pattern will be assigned to class +1, otherwise to class -1. In a case like this the decision boundary is a hyperplane in the N dimensional hyperspace. This linearity however has the disadvantage that the XOR problem is not solvable with a single-layer perceptron although it has a good generalization ability and the training algorithms are likely to converge to the optimal solution.

To solve the XOR problem for which it is not possible to classify all cases by a linear boundary additional hidden units have to be introduced which thereby extend the network to a *multi-layer* perceptron. These hidden nodes lie between the input and output layers (see Figure 3.4). They receive their inputs from a layer directly below and send their output to units in a layer directly above the units. There are no inter-layer connections. Generally multi-layer perceptrons consist of an arbitrary number of hidden layers with an arbitrary number of hidden nodes per layer. However it has been shown [66, 67, 68, 69] that only one layer of hidden units suffices to approximate any function with finitely many discontinuities to arbitrary precision, provided the activation functions of the hidden units are non-linear (the universal approximation theorem). In most applications a feed-forward network with a single layer of hidden units is used with a sigmoid activation function for the units leading to the following explicit expression for the complete function represented by the network:

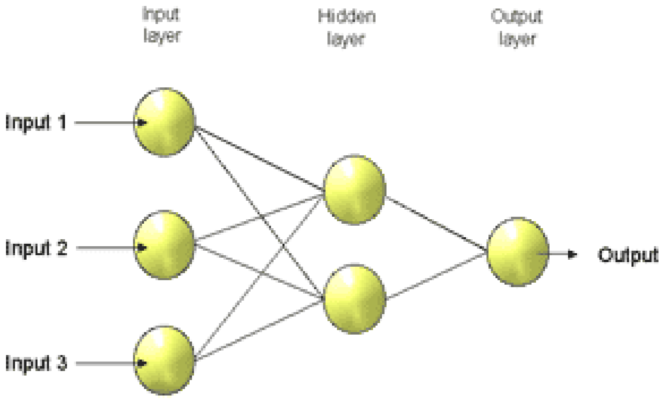

Figure 3.4: *Example of a multi-layer perceptron with one hidden layer.*

$$y = \mathcal{F}\left(\sum_{i=1}^{M} w_i^{(2)} \mathcal{F}\left(\sum_{j=1}^{N} w_{ij}^{(1)} x_j - \theta_i^{(1)}\right) - \theta^{(2)}\right) \tag{3.6}$$

where N denotes the number of input units and M the number of hidden units.

Practically the number of input and output units is defined by the problem. The number of hidden units must be adjusted to the special case under investigation. There is no general rule to determine this number.

3.1.4 Training a Neural Network

After describing the general architecture and functionality of a neural network in the previous Section this Section focuses on the training of such networks. The weights which form the free parameters of the neural network have to be configured such that the application of a set of inputs $\mathbf{x}^p$, where p denotes the pattern produces (either 'direct' or via relaxation process) the desired set of outputs $\mathbf{d}^p$ or target values. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using *a priori* knowledge. Another way is to 'train' the neural network by feeding it teaching patterns and letting it change its weights according to some learning rule.

Error Function

We first have to define an adequate error measure. All neural network training algorithms try to minimize the error of the set of learning samples which are available for training the network. The most commonly used error measure is the summed square error, as indicated by the least mean square.

$$E = \sum_p E^p = \frac{1}{2} \sum_p (d^p - y^p)^2 \tag{3.7}$$

where the index p ranges over the set of input patterns and E^p represents the error on pattern p. The choice of this error function is motivated by its analytical simplicity; several other error measures exist. In the case of two-class classification a more appropriate measure is the two-class (cross-)entropy error which was suggested by several authors in the late 1980's [70, 71, 72]:

$$E = \sum_p E^p = \frac{1}{2} \sum_p \left[(1 + d^p) \ln \frac{1 + d^p}{1 + y^p} + (1 - d^p) \ln \frac{1 - d^p}{1 - y^p} \right] \tag{3.8}$$

This error measure can be derived from the *relative entropy* of a probability function p_k with respect to a probability function q_k:

$$E = \sum_k p_k \ln \left(\frac{p_k}{q_k} \right) \tag{3.9}$$

This is motivated by the assumption that desired outputs d^p are independent, binary, random variables and the required network response represents the conditional probability that these variables would be one.

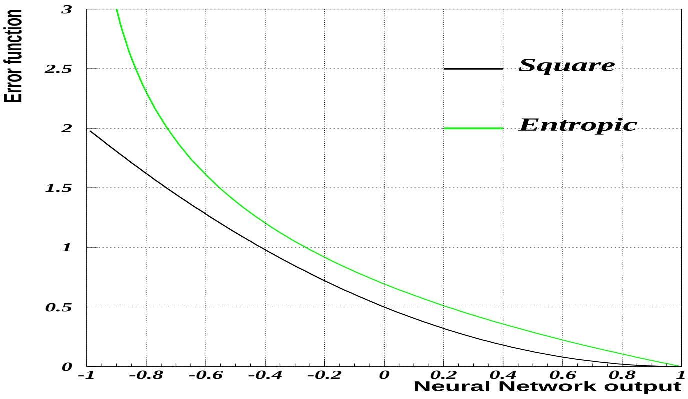

Figure 3.5: *Comparison of the square and the cross-entropic error function. The desired target output is $d = 1$. The cross-entropic error function is penalizing large deviations from the target value much more than the square error function.*

The cross-entropic error function of equation 3.8 requires the desired target to be $d = 1$ or $d = -1$. Figure 3.5 shows the two mentioned error functions in comparison. The advantage of the cross-entropy error function is the fact that the penalty for completely wrong classification goes to infinity.

Gradient Descent

The problem of finding suitable weights to solve a classification task is now translated into the problem of minimizing the error or cost function with respect to these weights. Given that the derivative of the non-linear transfer function is easy to evaluate which is the case for the logistic function, the least mean square procedure finds the values of all the weights that minimize the error function by a method called *gradient descent*. The idea is to make a change in the weight proportional to the negative of the derivative of the error as measured on the current pattern with respect to each weight:

$$\Delta \mathbf{w} = -\eta \nabla E(\mathbf{w}) \quad \eta > 0 \tag{3.10}$$

where η is a small positive number called the learning rate parameter. From

$$dE = \nabla E \cdot d\mathbf{w} \approx \nabla E \cdot \Delta \mathbf{w} = -\eta |\nabla E(\mathbf{w})|^2 < 0 \tag{3.11}$$

follows that the error function changes in fact to smaller values by gradient descent method. We assume that the activation is a differentiable function of the total input, given by

$$y_k^p = \mathcal{F}(s_k^p), \tag{3.12}$$

in which

$$s_k^p = \sum_j w_{jk} x_j^p + \theta_k \quad . \tag{3.13}$$

The error measure E^p is defined as the total quadratic error for pattern p at the output units k:

$$E^p = \frac{1}{2}\sum_k (d_k^p - y_k^p)^2 \tag{3.14}$$

where d_k^p is the desired output for unit k when pattern p is presented to the network. We further set $E = \sum_p E^p$. Applying the gradient descent method, we can write:

$$\frac{\partial E^p}{\partial w_{jk}} = \frac{\partial E^p}{\partial s_k^p}\frac{\partial s_k^p}{\partial w_{jk}} \quad . \tag{3.15}$$

By Equation 3.13 we see that the second factor is

$$\frac{\partial s_k^p}{\partial w_{jk}} = x_j^p \tag{3.16}$$

The derivatives of the error function for pattern p become

$$\frac{\partial E^p}{\partial s_k^p} = -\mathcal{F}'(s_k^p)\delta_k^p \tag{3.17}$$

with $\delta_k^p \equiv d_k^p - y_k^p$. Combining Equations 3.17 and 3.16 with Equation 3.10 the change in the weights due to presentation of a particular pattern is given by the so-called *delta rule*:

$$\Delta^p w_{jk} = \eta \mathcal{F}'(s_k^p)\delta_k^p x_j^p \tag{3.18}$$

where $\mathcal{F}'(s_k^p)$ can be expressed for the case of the logistic transfer function by a simple expression:

$$\mathcal{F}'(s_k^p) = \mathcal{F}(s_k^p)(1 - \mathcal{F}(s_k^p)) \tag{3.19}$$

All this can be generalized to an arbitrary number of patterns by computing the derivatives for each pattern separately and then summing over all patterns.

Back-Propagation

The most common training algorithm for feed-forward neural networks is the so-called *back-propagation* algorithm, which is based on the gradient descent method described above. Just using the gradient descent method without modification is not possible in the case of multi-layer networks. Applying the above mentioned delta rule would result in changing only the final layer of weights leaving the weights of previous layers unchanged. There would be no way to decide which weights in previous layers to change and by how much. This would result of course in loosing all advantages of multi-layer architectures mentioned in previous Sections.

The way out of this dilemma is simple. The error measure can be written as a function of the net inputs from hidden to output layer $E^p = E^p(s_h^p)$ and we use the chain rule to write

$$\frac{\partial E^p}{\partial s_h^p} = \frac{\partial E^p}{\partial y_h^p}\frac{\partial y_h^p}{\partial s_h^p} = \mathcal{F}'(s_h^p)\frac{\partial E^p}{\partial y_h^p} \tag{3.20}$$

with index h denoting the hidden node h. Calculating further leads to:

$$\frac{\partial E^p}{\partial y_h^p} = \sum_{o=1}^{N_o}\frac{\partial E^p}{\partial s_o^p}\frac{\partial s_o^p}{\partial y_h^p} = \sum_{o=1}^{N_o}\frac{\partial E^p}{\partial s_o^p}\frac{\partial}{\partial y_h^p}\sum_{j=1}^{N_h} w_{jo}y_j^p = \sum_{o=1}^{N_o}\frac{\partial E^p}{\partial s_o^p}w_{ho} = -\sum_{o=1}^{N_o} w_{ho}\mathcal{F}'(s_o^p)\delta_o^p \tag{3.21}$$

substituting this into equation 3.15 yields

$$\frac{\partial E^p}{\partial w_{jh}} = -\mathcal{F}'(s_h^p)\sum_{o=1}^{N_o} w_{ho}\mathcal{F}'(s_o^p)\delta_o^p\frac{\partial s_h^p}{\partial w_{jh}} \tag{3.22}$$

and

$$\Delta^p w_{jh} = \eta\mathcal{F}'(s_h^p)\sum_{o=1}^{N_o} w_{ho}\mathcal{F}'(s_o^p)\delta_o^p x_j^p \tag{3.23}$$

for the change in the weight between the input layer and hidden layer. Using the simple delta rule for the weights between the hidden layer and the output layer leads to

$$\Delta^p w_{ho} = \eta\mathcal{F}'(s_o^p)\delta_o^p y_h^p \tag{3.24}$$

To get the total change in the weight, everything has to be summed up over the number of patterns p presented to the network.

The equations derived in this section are mathematically correct, but there is a way of understanding back-propagation other than reciting the necessary equations. What happens in the above equations is the following. When a learning pattern is clamped, the activation values are propagated to the output units, and the actual network output is compared with the desired output values, we usually end up with an error in each of the output units. Let's call this error e_o for a particular output unit o. We have to bring e_o to zero.

The simplest way to do this quite simple: we strive to change the connections in the neural network in such a way that, next time around, the error e_o will be zero for this particular pattern. We know from the delta rule that, in order to reduce an error, we have to adapt its incoming weights according to

$$\Delta w_{ho} = \eta\mathcal{F}'(s_o)(d_o - y_o)Y_h \quad . \tag{3.25}$$

But this alone is not enough: when only applying this rule, the weights from input to hidden units are never changed, and we do not have the full representational power of the feed-forward network as promised by the universal approximation theorem. In order to adapt the weights from input to hidden units, we again want to apply the delta rule. in this case, however we do not have a value for δ for the hidden units. This is solved by the chain rule which does the following: distribute the error of an output unit o to all the hidden units that it is connected to, weighted by this connection. Differently put, a hidden unit h receives a delta from each output unit o equal to the delta of that output unit weighted with (=multiplied by) the weight of the connection between those units. In symbols: $\delta_h = \sum_o \delta_o w_{ho}$. Well, not exactly: we forgot the activation function of the hidden unit; $\mathcal{F}$ has to be applied to the delta before the back-propagation can continue.

3.1.5 Advanced Algorithms

In the previous section the basic techniques of neural network training have been described. Despite the apparent success of the back-propagation learning algorithm, there are some aspects which make the algorithm not guaranteed to be universally useful. Most troublesome is the long training process. This can be a result of a non-optimum learning rate and learning algorithm. Outright training failures generally arise from two sources: *network paralysis* and *local minima*.

- **Network paralysis**. As the network trains, the weights can be adjusted to very large values. The total input of a hidden unit or output unit can therefore reach very high (either positive or negative) values, and because of the sigmoid activation function the unit will have an activation very close to zero or very close to one. As is clear from Equation 3.19, the weight adjustments which are proportional to the first derivative of the activation function will be close to zero, and the training process can come to a virtual standstill.

- **Local minima**. The error surface of a complex network is full of hills and valleys. Because of the gradient descent, the network can get trapped in a local minimum, when there is a much deeper minimum nearby. Probabilistic methods can help to avoid this trap, but they tend to be slow. Another suggested possibility is to increase the number of hidden units. Although this will work because of the higher dimensionality of the error space, it appears that there is some upper limit of the number of hidden units which, when exceeded, again results in the system being trapped in local minima.

This section focuses therefore on some techniques to avoid such problems. This can be achieved in either preprocessing the input data of a neural network or by using more efficient algorithms. To improve the performance and the stabilization of network training an adequate mixture of both is preferable.

Learning Per Pattern

At each iteration the back-propagation algorithm requires a complete pass through the entire dataset to ensure the weights to be adjusted in order to decrease the error measure, in the direction of the *true* gradient. This is called *batch* learning since an entire batch of data must be considered before weights are updated. Alternatively one can use *stochastic* learning, where the learning rule is applied to each pattern separately, i.e., a pattern p is applied (preferably chosen randomly), E^p is calculated, and the weights are adapted. Because this method estimates the gradient by choosing one pattern, the weights may not move precisely down the gradient at each iteration. This *noisiness* at each iteration has the advantage that it makes it possible to get out of a local minimum by jumping into a neighboring, possibly deeper minimum. Another advantage of the stochastic learning is that there exists empirical indication, that it converges often much faster than the batch learning. This can be understood if one considers datasets with redundant information. Whereas in batch learning one loops over redundant pieces of information calculating several times the same quantity in stochastic learning only one iteration leads to the same result. Care has to be taken, however, with the order in which the patterns are taught. For example, when using the same sequence over and over again the network may become focused on the first few patterns. This problem can be overcome by using a permuted training method.

Learning Rate and Momentum

Training a Neural Network requires to choose an appropriate learning rate η. To ensure fast learning η has to be chosen as large as possible. If η is chosen too large however the weights will oscillate around the minimum and even worse it can lead to a situation, where the training process diverges. To avoid oscillation for high η one possibility is to make the change in weight dependent of the past weight change by adding a *momentum* term:

$$\Delta\mathbf{w}(t+1) = -\eta\nabla E(\mathbf{w}) + \alpha\Delta\mathbf{w}(t) \quad , \tag{3.26}$$

where t indexes the presentation number and α is a constant which determines the effect of the previous weight change.

Figure 3.6 illustrates the learning process for different learning rates η and how important an effective choice is.

Conjugate Gradient

There exist many improvements for the back-propagation algorithm described in this Chapter. Maybe the most obvious improvement is to replace the rather

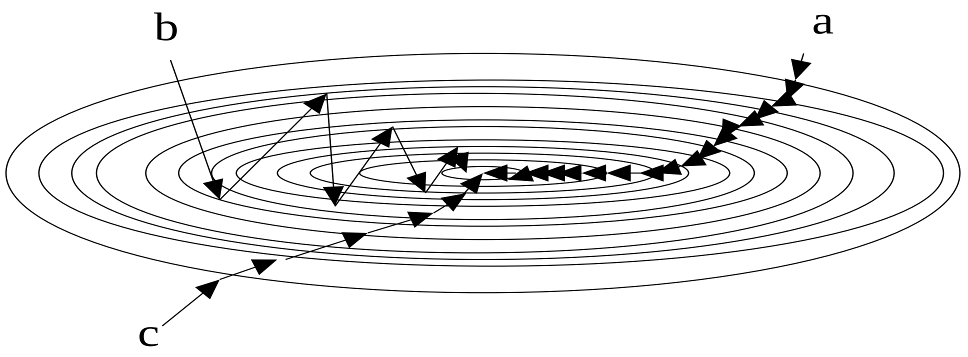

Figure 3.6: *The descent in weight space. a) for small learning rate; b) for large learning rate: note the oscillations, and c) with large learning rate and momentum term added.*

primitive steepest descent method with a *direction set* minimization method, e.g., *conjugate gradient minimization.* Assume we pick a descent direction, e.g. the gradient, then we minimize along a line in this direction (line search). This brings the error function E at a place where its gradient is perpendicular to the primary direction (otherwise minimization along this direction is not complete). Subsequently we should try to find a direction along which the gradient does not change its direction, but merely its length (conjugate direction, such that minimization along this new direction does not spoil the result of the previous minimization, i.e., the directions are non-interfering.The evolution of the descent directions p_k at iteration k is given as

$$p_k = -\nabla E(\mathbf{w}_k) + \beta_k p_{k-1} \quad , \tag{3.27}$$

where the choice of β_k can be done either according to Fletcher and Reeves [73]

$$\beta_k = \frac{\nabla E(\mathbf{w}_k)^T \nabla E(\mathbf{w}_k)}{\nabla E(\mathbf{w}_{k-1})^T \nabla E(\mathbf{w}_{k-1})} \tag{3.28}$$

or Polak and Ribiere [74]

$$\beta_k = \frac{(\nabla E(\mathbf{w}_k) - \nabla E(\mathbf{w}_{k-1}))^T \nabla E(\mathbf{w}_k)}{\nabla E(\mathbf{w}_{k-1})^T \nabla E(\mathbf{w}_{k-1})} \tag{3.29}$$

Two directions p_k and p_{k-1} are defined as conjugate if

$$p_k^T H p_{k-1} = 0 \tag{3.30}$$

where H denotes the Hessian matrix whose components are given by

$$H_{ij} = \frac{\partial^2 E}{\partial w_i \partial w_j} \tag{3.31}$$

For both choices it is very important to have a good line search procedure. For a perfectly quadratic function with N variables a convergence within N steps can be proven.

Weight Decay and Pruning

Weight decay adds a penalty term to the error function. The usual penalty is the sum of squared weights times a decay constant c leading to the modified error function:

$$E(\mathbf{w}) \to \tilde{E}(\mathbf{w}) = E(\mathbf{w}) + \frac{c}{2}\sum_i w_i^2 \tag{3.32}$$

If the data term is small compared to the weight decay term, the weight update rule leads to

$$\begin{aligned} \Delta w_i &= -\eta c w_i \\ \longrightarrow w_i(t+1) &= w_i(t)(1-\eta c) \\ \longrightarrow w_i(t) &= w_i(0)(1-\eta c)^t = w_i(0)e^{t\ln(1-\eta c)} \end{aligned} \tag{3.33}$$

where t indexes the iteration number. This represents the exponential decay of the weights to zero if $0 < \eta \cdot c < 1$. Weight decay is a subset of regularization methods which are very useful to get control over the network complexity and to deal with ill-conditioned non-linear least square problems leading to oscillatory behavior of the network. The penalty term in weight decay, by definition, penalizes large weights. It causes the weights to converge to smaller absolute values than they otherwise would. Large weights can hurt generalization in two different ways. Excessively large weights leading to hidden units can cause the output function to be too rough, possibly with near discontinuities. Excessively large weights leading to output units can cause wild outputs far beyond the range of the data if the output activation function is not bounded to the same range as the data. To put it another way, large weights can cause excessive variance of the output (see [75]).

Other penalty terms besides the sum of squared weights are sometimes used. Weight elimination (see [76]) uses:

$$\sum_i \frac{w_i^2}{w_i^2 + c^2} \tag{3.34}$$

where w_i is the ith weight and c is a user-specified constant. Whereas decay using the sum of squared weights tends to shrink the large coefficients more than the small ones, weight elimination tends to shrink the small coefficients more, and is therefore more useful for suggesting subset models (*pruning*). Pruning techniques set weights to zero and thus remove them from the network topology if the value of a weight falls below a certain threshold. If all connections of a certain node are removed in this way, the entire node is removed. In this way the network is able to optimize its own topology since only the nodes and connections between the nodes that the network regards as important are kept.

3.1.6 Data Preprocessing

In order to make the training faster and to enhance the probability that it converges one can improve a few things in the way data is presented to the network. To ensure that the network does only learn those features from the presented data which are important for the classification task it is useful to transform the data before feeding them into the neural network. This transformation can be done by hand using physics knowledge or by using automatized methods which do a pre-processing based on the data distributions of the input variables. This preprocessing is able to improve the pattern recognition system by giving the network inputs which are simplified and easy to handle and therefore a better minimum can be found. Some of these preprocessing ideas and why they work are described in this Section.

Choosing the Inputs

A Neural Network is in general a tool which is quite good in generalizing and finding separating non-linear hypersurfaces in multidimensional input space, however, it is far from being capable in handling an arbitrary amount of input variables. Each input variable used adds one dimension to the problem which means that a minimum has to be found in a space which is one dimension higher. Not only the convergence properties might be poor, also computing time increases drastically using a lot of different input variables.

A better way is to understand which variables give a useful contribution to the information needed and which variables are redundant or useless. There might be cases, however, where it is useful to include variables which have nearly no correlation to the desired output but are in correlation with other useful variables. For example if one includes so-called *quality* variables which represent estimates for how useful the other variables might be concerning the classification task, this gives the neural network quite useful information on how to treat information. If an input sample is clearly belonging to one class, but the quality is so poor, that it has nearly no meaning, the Neural Network will learn this feature and will not take much care of this sample than without giving it the quality variable.

Choosing variables properly means, that one also has to take into account that the training data might differ in some systematic way from the actual data, e.g. when the function being modeled is changing over time, and so the relationships which held in the past may no longer hold. Also including variables which are correlated with variables under investigation is not very useful because one might get unwanted biases and artifacts.

For stochastic learning another consideration is important. The order in which the training samples are presented to the network are important. To increase the learning speed, it is advisable to choose a sample at each iteration that is the most unfamiliar to the system. To implement this one can choose successive examples that are from different classes since training examples belonging to the same class will most likely contain similar information.

Transforming the Inputs

After selecting the important variables from the unimportant variables it is advisable to apply further treatment to the variables themselves. Convergence is usually faster if the average of each input variable over the training set is close to zero. One can understand this behavior if one considers the extreme case where all the inputs are positive. This would result in the updates of the weights in the first weight layer having all the same sign. As a result , these weights can only all decrease or increase *together*. Thus, if a weight vector must change direction it can only do so by zigzagging which is inefficient and thus very slow. For this reason it is good to shift the inputs so that the average over the training set is close to zero. This however cannot be applied for the outputs of the hidden nodes which are inputs for the next layer. Instead it is advisable to use a sigmoid function which is centered around zero.

Learning is also faster if the input variables are scaled so that all have about the same covariance. The reason for this is that the learning speed is balanced out between the different input variables since generally the learning rate η is chosen globally for all weight updates. It is also advisable to match the covariance of the inputs with that of the transfer function used.

Another quite useful preprocessing step is the decorrelation of input variables. If inputs are highly correlated the training is much slower because the network has to deal with degeneracies which means flat areas on the error surface. In the worst case two or more input variables are linear dependent, and so the network is fed with information which does not improve learning and therefore it is unimportant. In order to decorrelate the variables their covariance matrix has to be diagonalized.

The above mentioned preprocessing techniques are illustrated in Figure 3.7.

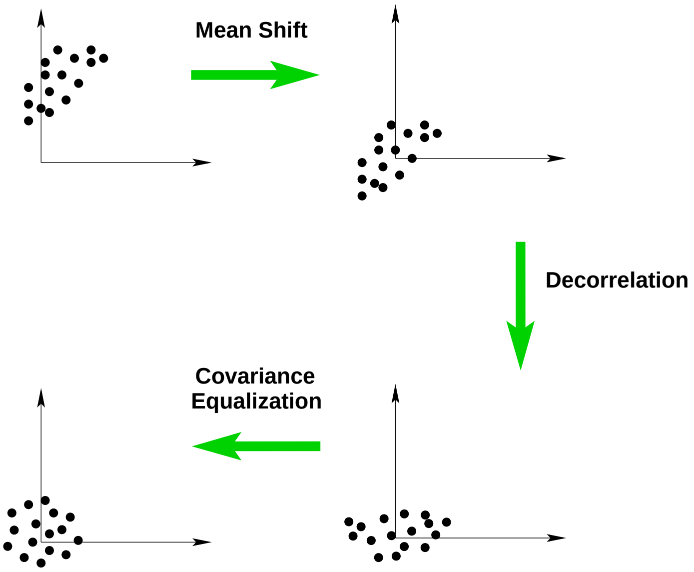

Figure 3.7: *The plots illustrate the subsequent transformation of the variables in the input space.*

3.1.7 Alternative Approaches

Neural Networks that are based on processing nodes described in Section 3.1.2 which rely on dot products and sigmoids can in principle approximate any functional mapping to arbitrary accuracy (universal approximation theorem), and therefore constitute a universal class of parametrised multivariate non-linear mappings. Although most systems are based on such architectures many other types of units (or layers) can be used. In this Section two common alternatives are described, namely the *Radial Basis Function* (RBF) (see [77, 78, 84, 85]) and the *Support Vector Machines* (SVM).

In the Karlsruhe DELPHI group both approaches have been tested and compared to the performance of multi-layer Neural Networks on classification problems for which the Neural Network approach already delivered very powerful classifiers. It turned out that for applications in high energy physics both alternatives manage to achieve a performance as good as Neural Networks. The big drawbacks however were the high number of parameters to be tuned in order to have the training converge fast and the huge amount of required CPU time (especially for support Vector Machines).

Radial Basis Function (RBF)

In RBF networks, the dot product of the weight and the input vector is replaced with a Euclidean distance between the input vector and the weight vector, and the sigmoid is replaced by a Gaussian function. The output activity is computed, e.g. for one output, as

$$\mathcal{F} = \sum_{i=1}^{N} \exp(-\frac{1}{2\sigma_i^2}||\mathbf{x} - \mathbf{w}_i||^2) \tag{3.35}$$

where N denotes the number of radial basis functions and $\mathbf{w}_i(\sigma_i)$ is the corresponding mean vector(standard deviation) of the RBF's which have to be adjusted. The norm or distance metric of course does not have to be Euclidean and can be chosen in whatever way seems most appropriate to the states and task at hand.

These RBF units can replace or coexist with the standard units and they are usually trained by a combination of gradient descent (for output units) and unsupervised clustering for determining the means and widths of the RBF units.

The motivation for using RBF's lies in the time needed for training which can be substantially faster than the methods used to train multi-layer perceptron networks. On the negative side, the locality property of RBF's may be a disadvantage particularly in high dimensional spaces because many units are needed to cover the spaces and noisy data results typically in high oscillatory functions which is generally undesirable.

Support Vector Machines (SVM)

Support Vector Machines are used for several pattern recognition problems like object recognition, image recognition, speaker identification and so on. The subject started in the late seventies [79] and since then a few steps are done to introduce this method in modern particle physics.

The method is based on the mathematical problem to find the best dividing hyperplane between two classes in a space of arbitrary dimension. The term *best* means that the distance between any representative of the classes and the hyperplane is maximal. To maximize the distance (the so-called margin) between the data points and the plane, which is $2/||\mathbf{w}||$, where $\mathbf{w}$ is the normal vector of the plane, one has to minimize $||\mathbf{w}||^2$ with the following constraints:

$$\begin{array}{lll} & & \text{vector } \mathbf{x}_i \text{ belongs to} \\ \mathbf{x}_i \cdot \mathbf{w} + b \geq +1 & \text{for } y_i = +1 & \text{(class 1, signal)} \qquad (3.36) \\ \mathbf{x}_i \cdot \mathbf{w} + b \leq -1 & \text{for } y_i = -1 & \text{(class 2, background)} \qquad (3.37) \end{array}$$

These conditions can be combined into one set of inequalities:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \quad \forall i \tag{3.38}$$

To solve this minimization one switches to a Lagrangian formulation of the problem. Positive Lagrange multipliers α_i, $i = 1, \dots, l$ are introduced, one for each of the inequality constraints 3.38. This results in the Lagrangian:

$$L \equiv \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{l} \alpha_i(y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1) \tag{3.39}$$

The minimum of $||\mathbf{w}||^2$ taking into account the inequality constraints is obtained when the partial derivatives of L with respect to $\mathbf{w}$ and b simultaneously and the derivatives with respect to the α_i vanish.

This leads to the following equations, the so-called *Karush-Kuhn-Tucker* (KKT) conditions [80]:

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{w}_\nu} L = \mathbf{w}_\nu - \sum_i \alpha_i y_i \mathbf{x}_{i\nu} &= 0 \quad \nu = 1, \dots, d && (3.40)\\
\frac{\partial}{\partial b} L = -\sum_i \alpha_i y_i &= 0 && (3.41)\\
y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 &\geq 0 \quad i = 1, ..., l && (3.42)\\
\alpha_i &\geq 0 \qquad \forall i && (3.43)\\
\alpha_i(y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1) &= 0 \qquad \forall i && (3.44)
\end{aligned}$$

Here d is the dimensionality of the feature space and l is the number of training vectors. The last Eqn. 3.44 follows from the fact that for vectors that fulfill the equality sign of relation (3.38), the constraint is active and the Lagrange multiplier is positive. The other vectors play no role, their corresponding constraints are inactive, thus their corresponding Lagrange multipliers α_i are zero. This is the mathematical formulation of the fact that only the input vectors which lie nearest to the plane (*Support Vectors*) are relevant. Leaving out the data points with α_i=0 would lead to the same solution.

For more convenience the problem is reformulated in the following. The main advantage of this so-called *Wolfe Dual* problem is that the data only appear in the form of dot products $\mathbf{x}_i \cdot \mathbf{x}_j$ which can be generalized for the treatment of non-liner mappings. The requirement that the gradient of L with respect to $\mathbf{w}$ and b vanishes (Eqns. 3.40 and 3.41) allows to express $\mathbf{w}$ and b as functions of α_i. These expressions can be substituted into Equation 3.39 to give

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \mathbf{x}_j) \tag{3.45}$$

This equation describes the Lagrangian in the minimum valley with respect to $\mathbf{w}$ and b, as a function of the α_i. We still have to search the maximum with respect to the α_i. It is a dual formulation of the same problem: instead of minimizing $||\mathbf{w}||^2$ with the inequality constraints 3.38, one can maximize this *dual* Lagrangian (3.45) with respect to the α_i.

All this is just valid for linear and separable problems, and only then we can obtain a solution. But in modern particle physics one usually has to handle non-linear, non-separable classification problems.

If one has to deal with non-linear problems, the idea is to transform the data points from the origin space into another space, where the problem hopefully becomes linear. This can be achieved by a mapping function Φ:

$$\Phi : \mathcal{R}^d \mapsto \mathcal{H} \tag{3.46}$$

Notice then that the only way in which the data appear is in the form of dot products $\mathbf{x}_i \cdot \mathbf{x}_j$. So the whole training algorithm would only depend on the data through dot products in $\mathcal{H}$, i.e. on functions of the form $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$. Now if there were a *kernel function* K such that $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$, one would only need to use K in the training algorithm, and would never need to explicitly even know what Φ is. Suppose that the data are vectors in $\mathcal{R}^2$, and we choose $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^2$. Then it is easy to find a space $\mathcal{H}$, and mapping Φ from $\mathcal{R}^2$ to $\mathcal{H}$, such that $(x \cdot y)^2 = \Phi(x) \cdot \Phi(y)$:

$$\Phi(x) = \begin{pmatrix} \mathbf{x}_1^2 \\ \sqrt{2}\mathbf{x}_1\mathbf{x}_2 \\ \mathbf{x}_2^2 \end{pmatrix} \tag{3.47}$$

Another example with large practical significance is the Gaussian kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-||\mathbf{x}_i - \mathbf{x}_j||^2/2\sigma^2}. \tag{3.48}$$

In this particular example, the dimension of $\mathcal{H}$ is infinite, so it would not be very easy to work with Φ explicitly. But that does not matter, since one just has to replace in the algorithm $\mathbf{x}_i \cdot \mathbf{x}_j$ by $K(\mathbf{x}_i, \mathbf{x}_j)$. All former considerations hold, since one is still performing a linear separation, but in a different space.

The Gaussian kernel is very important since it allows for constructing local clusters in space. Its applications can be compared to Radial Basis Function Neural Networks described in the previous Section. In modern particle physics one mostly has to deal with non-separable data sets, i.e. a statistical separation between signal and background can be achieved, but there are always single background events looking like signal. In such cases the above algorithm will find no feasible solution, since it has to find a classification which is 100% correct.

To apply SVM also in these non-separable cases, one would like to relax the constraints 3.38, but only when necessary. This is achieved by extending the above ideas by introducing positive slack variables $\xi_i, i = 1, \dots l$ in the constraints [81], which then become:

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq +1 - \xi_i \quad for\ y_i = +1 \tag{3.49}$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq -1 + \xi_i \quad for\ y_i = -1 \tag{3.50}$$

$$\xi_i \geq 0 \qquad \forall i \tag{3.51}$$

Thus, for a classification error to occur, the corresponding ξ_i must exceed unity, so $\sum_i \xi_i$ is an upper bound on the number of training errors. Therefore the Lagrangian has to be changed from $\|\mathbf{w}\|^2/2$ to $\|\mathbf{w}\|^2/2 + C(\sum_i \xi_i)$, where C is a parameter to be chosen by the user, a large C corresponding to assigning a higher penalty to errors. So C is a measure for the softness of the margin. Then the dual problem becomes:

Maximize:

$$L_D \equiv \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \tag{3.52}$$

subject to the constraints:

$$0 \leq \alpha_i \leq C, \tag{3.53}$$

$$\sum_i \alpha_i y_i = 0. \tag{3.54}$$

The solution is again given by

$$\mathbf{w} = \sum_i^{N_s} \alpha_i y_i \mathbf{x}_i \tag{3.55}$$

where N_s is the number of Support Vectors. Thus the only difference from the optimal hyperplane case is that the α_i now have an upper bound of C. The situation is summarized schematically in Figure 3.8.

3.1.8 Dedicated Problems

Neural Network techniques are applicable to an enormous amount of tasks for which the setup for the network has to be made individually. They can be divided into mainly three classes of applications, namely *classification*, *regression* and *probability density function*. This Section focuses on these application classes with respect to the importance in high energy physics especially in this analysis.

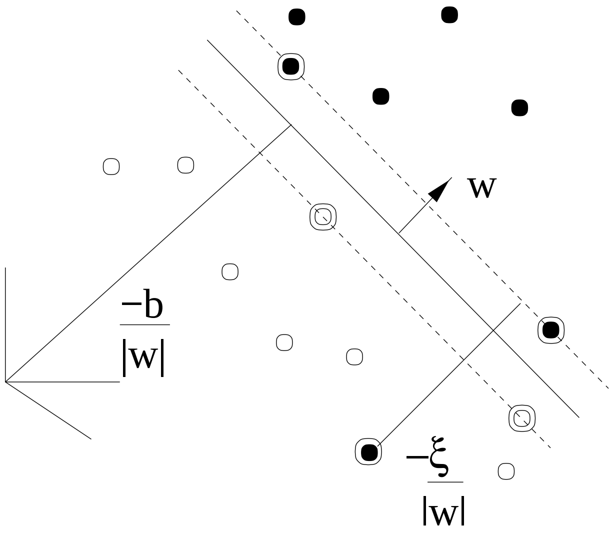

Figure 3.8: *Linear separating hyperplanes for non-separable case. Support Vectors are marked with circles. Effectively the data points inside the margin are not classified as classification errors. [82]*

Classification

The general objective in classification applications of neural networks is to determine to which of a number of discrete classes a given input case belongs. As mentioned earlier a useful interpretation of the network output was as estimate of the probability of class membership, in which case the network was actually learning to estimate a probability density function (p.d.f.). This probability density function is the conditioned probability $P(\mathcal{C}_k|\mathbf{x})$ of input vector $\mathbf{x}$ belonging to class $\mathcal{C}_k$. This interpretation however relies on certain conditions. These conditions are:

- The output of the Neural Network has to lie between 0 and 1. Therefore the non-linear activation function has to be the logistic sigmoid function.
- The error function has to be either the quadratic or the logarithmic function based on entropy
- The nodes have to have standard Gaussian distributed inputs.

With these requirements met the output of such a probabilistic Neural Network may be interpreted directly as a probability density function.

Regression

In regression problems the purpose is to predict the value of a continuous output variable. It can be used to approximate a continuous function or to predict a value in the future from a number of time-ordered values in the past. From the universal approximation theorem follows that in order to approximate any continuous function a Neural Network with one hidden layer is sufficient.

Probability Density Reconstruction

In this application the aim is not to predict a single value of a function or a probability density function like in classification or regression, but to deliver the complete probability density function of the output variable. From this function estimates for the mean, the median, the maximum likelihood, the uncertainty or the moments can be deduced. To have a Bayesian interpretation of the estimator one has to take into account a priori knowledge in the form of the inclusive (unconditional) distribution. This way the prediction will never result in unphysical values outside the training range. A quite powerful approach of a neural network for probability density reconstruction has been developed in Karlsruhe [83].

The probability density reconstruction can technically be realized by solving binary classification problems for various output nodes. Namely the classification problem for each output node is *the true value is above the threshold value for this node* vs. *the true value is below the threshold value for this node* where the true value t is transformed to its cumulative probability density function s.

$$s = F(t) = \int\limits_{t_{min}}^{t} f(t')\,dt' \tag{3.56}$$

Therefore the desired reconstructed probability density function is the derivative of the network outputs.

As an example, the desired target output vector of an event with s_{true} and 10 output nodes would be $\vec{T} = (+1, +1, +1, +1, +1, +1, -1, -1, -1, -1)$. Note that $s_{true} = 0.56$ and $s_{true} = 0.64$ would give the same output vector. This is the discretization uncertainty introduced. Therefore an important consideration is the number of output nodes since this limits the best resolution the Neural Network can achieve. To reduce the discretization information loss the target node nearest to the true value s_{true} can be set to the value between $+1$ and -1 to smoothen the dependence. In this example the target vector would be $\vec{T} = (+1, +1, +1, +1, +1, +0.2, -1, -1, -1, -1)$ for $s_{true} = 0.56$ but $\vec{T} = (+1, +1, +1, +1, +1, +1, -0.2, -1, -1, -1)$ for $s_{true} = 0.64$.

The output points of the neural network is then fitted by a cubic B-spline in order to provide a parametric form of the network output. In order to achieve smooth functions the fit is regularized on the basis of the sum of the third derivatives squared.

3.1.9 Application in High Energy Physics

The above described Neural Network techniques have in recent years been used increasingly in high energy physics. More and more physics analyses rely on Neural Networks as classifiers on a basis of Bayesian interpretation. Starting with classifiers on the track level Neural Network helped to improve traditional approaches in particle identification [102].

Especially for this analysis classification and probability density reconstruction applications were quite important for b-hadron energy reconstruction and for the enhancement of different b-hadron species which will be described in the next Chapter. Also for other analysis like the reconstruction of excited b mesons which uses a dedicated Neural Networks to improve the b-hadron flight direction resolution or oscillation analysis which relies on flavor tagging, neural network techniques are proven to be quite useful and successful.

3.2 Monte Carlo Integration

In high energy physics a common problem is related to limitations of a measurement of a certain physics quantity, e.g. b-hadron decay length, or b-hadron momentum. These limitations arise from detector inefficiencies or from undetected tracks. From simulated events one can deduce the dependency of the reconstructed quantity from the true quantities. The effect of the imperfect information from a real detector can be extracted in terms of *resolution functions*. By convoluting a theoretical a priori prediction of the true values of a physics quantity with these resolution functions it is possible to get an estimate of the expected measured distribution:

$$f(x_{rec}) = f(x_{true}) \otimes R(x_{rec}|x_{true}) = \int f(x_{true}) \cdot R(x_{rec}|x_{true})\, dx_{true} \quad , \qquad (3.57)$$

which can be deduced by the Bayes' theorem considering the resolution $R(x_{rec}|x_{true})$ as a conditional probability density of measuring x_{rec} if the true value is x_{true}.

In general these integrals are not solvable analytically and so one has to realize them by numerical integration techniques. In this analysis however the situation is quite difficult because the measurement of the decay time of a b-hadron requires the measurement of two quantities, the b-hadron momentum and the b-hadron decay length. Therefore providing an estimate of the expected distribution requires a two-dimensional integration event by event. Using a large amount of statistics leads to the fact that computing power becomes a limiting factor using standard Monte Carlo integration techniques. In order to enhance integration speed two approaches have been developed and find their application in this analysis.

The first approach addresses the point set used for the integration. The integration error, defined as the difference between numerical integration and analytical integration depends strongly on the number N of support points used for the integration. In order to diminish computation time one would like to have a point set with fewest possible support points while still having an acceptable accuracy. Using a normal grid as point set the error drops with $1/N$ in two dimensions. Using normal random numbers instead, it can be shown, that the error converges with $1/\sqrt{N}$. However, the best solution for this application is a point set based on so-called *Quasi Random Numbers* where the error converges faster compared to points on a grid and which are also advantageous compared to random numbers. Section 3.2.1 gives a brief introduction to quasi random number algorithms.

The integration speed can be further enhanced by varying the point density according to the integrand. One would like to enhance the point density in regions were the integrand has a high curvature. On the other hand, in regions were the integrand is constant or changes only linearly only a few support points are necessary to get an arbitrary integration precision.This method is called *importance sampling*. It is realized in form of an analytic variable transformation, made possible by the fact, that the shape of the resolution functions in this analysis can be modeled either by Gaussians (for the b momentum) or by modified Breit-Wigner functions (for the b decay length) and therefore have only one region where the point density has to be increased. The importance sampling is described in detail in Section 3.2.2.

3.2.1 Quasi Random Numbers

General Monte Carlo Techniques

The problem to solve is the numerical calculation of a multidimensional integral:

$$I = \int_B f(\mathbf{x})\, d\mathbf{x} \tag{3.58}$$

where $\mathbf{x} = x^\mu = (x^1, x^2, \ldots, x^s)$ denotes a point in the s-dimensional integration region B. The above integral can be expressed in a different form by:

$$I = vol(B) \int_B vol(B)^{-1} c_B(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} = vol(B) \langle f(\mathbf{x}) \rangle \tag{3.59}$$

Here $\langle f(\mathbf{x})\rangle$ is the expectation value of the function f for the uniform probability distribution $u(\mathbf{x}) = vol(B)^{-1}c_B$ which is defined in $B \subseteq \mathcal{R}^2$ where c_B is the characteristic function in the integration region. In the one-dimensional case with the integration boundaries a and b this gives:

$$u(x) = \begin{cases} \frac{1}{b-a} \cdot 1 & : \quad x \in [a,b] \\ \frac{1}{b-a} \cdot 0 & : \quad x \notin [a,b] \end{cases} \tag{3.60}$$

And the integral becomes:

$$I = (b-a)\langle f(x)\rangle \tag{3.61}$$

A fundamental statistical method is to extract properties of random variables like the standard deviation or mean value by looking at random samples of this variable. Therefore one can estimate the expectation value of $f(x)$ by selecting randomly distributed support points x_i and calculating the mean of $f(x_i)$:

$$\langle f(x)\rangle \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i) \equiv \frac{S}{b-a} \tag{3.62}$$

The integration error using this estimate is

$$\epsilon = I - S \tag{3.63}$$

From the Central Limit Theorem one can derive the following properties for a sample of random points which are distributed identically and independently: The error ϵ is normally distributed around zero and the standard deviation scales with $\sqrt{V/N}$ where V denotes a measure of the variation of the integrand (the proof of this can be found e.g. in [86]). This Monte Carlo method shows a convergent behavior which is independent of the dimension s. Using a grid instead, the error ϵ drops with $1/(\sqrt[s]{N})^2$. Therefore Monte Carlo integration is better if

$$\frac{1}{2} \geq \frac{2}{s} \longrightarrow s \geq 4 \quad . \tag{3.64}$$

The task now is to find a point set showing a more convergent behavior.

Koksma-Hlawka Inequality Equation

The *Koksma-Hlawka Inequality Equation* delivers a boundary of the integration error mentioned in the previous Section. For uniformly distributed sequences of numbers $P(N)$ the error boundary is:

$$\epsilon = |I - S| \leq V(f)D^*_\infty(P_N) \tag{3.65}$$

where $V(f)$ denotes the variation of the function f defined by Hardy and Krause in [87] and $D^*_\infty(P_N)$ the discrepancy of the sequence $P(N)$ which is, roughly speaking, the deviation from a perfect uniform distribution. The proof of the Koksma-Hlawka inequality equation can be found in [88].

Discrepancy

Discrepancy is a measure of the uniformity of a set of points. There exist various definitions of discrepancies using the deviation from a uniform distribution. Assuming a volume rectangle K with volume $vol(K)$ inside an s dimensional space I^s with volume 1 with sides parallel to the coordinate axis one can approximate this volume using a point set P_N in this space and counting the numbers of points $A_P(K;N)$ inside K divided by the number of points N. The discrepancy mentioned in the previous section is then defined as:

$$D^*_\infty(P_N) = \sup_{K \in I^s} |R_N(K)| = \sup_{K \in I^s} \left| \frac{A_P(K;N)}{N} - vol(K) \right| \tag{3.66}$$

By the law of iterated logarithms, the expectation value of the discrepancy of a random sequence is bounded by $(\log\log N)N^{-1/2}$. There are many quasi random sequences known for which the discrepancy is bounded by a constant times $(\log N)^s/N$, which suggests greater uniformity than a random sequence and therefore faster convergence in numerical integration.

A convenient measure of quality for a certain point set is the quadratic discrepancy normalized to the quadratic discrepancy for truly random numbers, called the normalized quadratic discrepancy, which is defined as:

$$\frac{(D^*_2(P_N))^2}{\langle (D^*_2(P_N))^2 \rangle} = \frac{(D^*_2(P_N))^2}{\frac{1}{N}\left(\frac{1}{2^s} - \frac{1}{3^s}\right)} \quad , \tag{3.67}$$

where $D^*_2(P_N)$ is defined as

$$D^*_2(P_N) = \left(\int\limits_{I^s} (|R_N(K)|)^2 \, dx \right)^2 \tag{3.68}$$

Niederreiter and Hammersly Generator

In Karlsruhe a dedicated analysis was made where different low discrepancy sequences were tested and their performance compared (see [89]). Based on these results the Hammersly method was chosen [89], which is a mixture of a linear grid in one dimension and the Niederreiter generator [90]. This Section gives a brief description of the generator.

Starting with the Halton sequence [91] in one dimension by choosing a prime b and expanding the sequence of integers $0, 1, 2, \ldots, N$ into base b notation the nth term of the sequence is given by

$$z_b(n) = \frac{a_0}{p} + \frac{a_1}{p^2} + \frac{a_2}{p^3} + \cdots + \frac{a_m}{p^{m+1}} \quad , \tag{3.69}$$

where the a_i's are integers taken from the base b expansion of $n-1$

$$[n-1]_b = a_m a_{m-1} \cdots a_2 a_1 a_0 \qquad \text{where} \quad m = m(n) \quad , \tag{3.70}$$

with $0 \leq a_i \leq b$. For example, if $b = 3$, the first 12 terms of the sequence ($n = 1 \ldots 12$) are

$$\{0, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}, \frac{19}{27}\} \tag{3.71}$$

To generalize this in s dimensions s one-dimensional Halton sequences based on s different primes are paired. Usually the first s prime numbers are chosen.

The Niederreiter generator is an improvement of the Halton generator by using additional functions $m_\mu(k)$, $\mu = 1, 2, \ldots, s$ and defining a new sequence

$$x_k^\mu = z_b(m_\mu(k)) \tag{3.72}$$

The functions m_μ^k are chosen in such a way that when $N = b^m$, the numbers $m_\mu(1), m_\mu(2), \ldots, m_\mu(N)$ are just a permutation of $(1, 2, \ldots, N)$.

The Niederreiter generator is the optimal choice for a wide range of numbers N. In special cases, for certain values of N a slightly better discrepancy can be obtained. The Hammersly algorithm is constructed by a grid in the first dimension ($x_n^1 = \frac{n-1}{N}$) and Niederreiter sequences for all the other dimensions. Figure 3.9 shows a comparison of the Hammersly generator in two dimensions to a normal pseudo random number generator. Clearly visible is the clustering of the pseudo random generator which increases the discrepancy.

Table 3.1 illustrates the performance of the Halton, Niederreiter and Hammersly generator on the basis of the normalized quadratic discrepancy. The chosen dimension is $s = 2$ and leads to an optimal point set for $N = 2^l$. The numbers are taken from [89].

N	Halton	Niederreiter D_2	Hammersly D_2
512	$1.119 \cdot 10^{-2}$	$5.815 \cdot 10^{-3}$	$4.844 \cdot 10^{-3}$
1024	$1.119 \cdot 10^{-2}$	$3.136 \cdot 10^{-3}$	$2.647 \cdot 10^{-3}$
2048	$5.544 \cdot 10^{-3}$	$1.896 \cdot 10^{-3}$	$1.453 \cdot 10^{-3}$
4096	$4.066 \cdot 10^{-3}$	$9.888 \cdot 10^{-4}$	$7.807 \cdot 10^{-4}$
8192	$2.076 \cdot 10^{-3}$	$5.493 \cdot 10^{-4}$	$4.220 \cdot 10^{-5}$

Table 3.1: *The normalized quadratic discrepancies for the Halton, Niederreiter and the Hammersly generator. The discrepancy is calculated for dimension* $s = 2$ *and leads to optimal point sets for* $N = 2^l$.

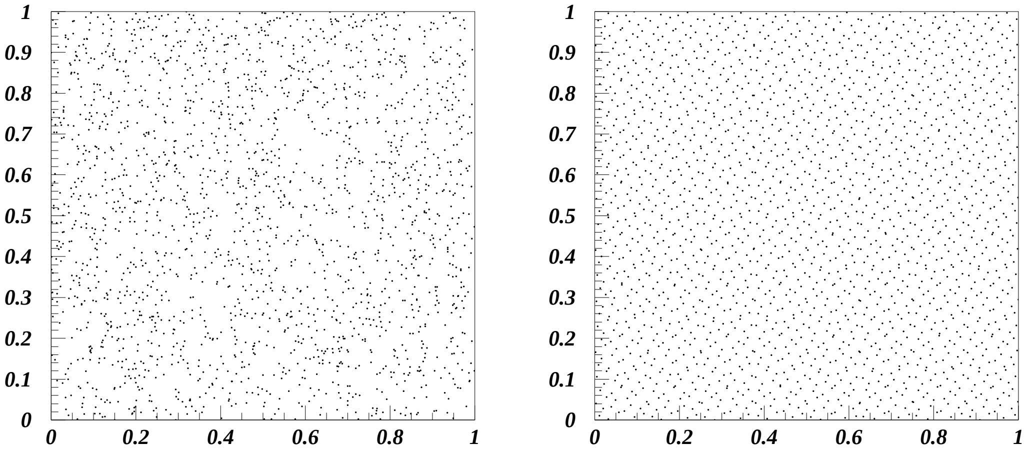

Figure 3.9: *The left plot illustrates a two-dimensional distribution of 2048 Pseudo Random points, the right plots shows 2048 points based on the Hammersly-Niederreiter algorithm. Clearly visible is the clustering of points for the pseudo random case which leads to high discrepancies.*

3.2.2 Importance Sampling

From the Koksma-Hlawka inequality equation 3.65 introduced in the previous Section one can learn, that the reduction of the integration error can be achieved either by using low discrepancy point sets or by reducing the variance of the integrand. As already stated in the introduction, the chosen technique is *Importance Sampling* which is realized by an analytical variable transformation as the shape of the integrand is basically known and can be mapped by an estimate of a central value and a width.

The problem of solving an integral (Equation 3.58) can be rewritten by introducing a probability density function $w(\mathbf{x})$:

$$I = \int\limits_B f(\mathbf{x})\, d\mathbf{x} = \int\limits_B \left[\frac{f(\mathbf{x})}{w(\mathbf{x})}\right] w(\mathbf{x})\, d\mathbf{x} = \left\langle \left[\frac{f(\mathbf{x})}{w(\mathbf{x})}\right] \right\rangle \tag{3.73}$$

with

$$\int\limits_B w(\mathbf{x})\, d\mathbf{x} = 1 \tag{3.74}$$

A Monte Carlo estimate of the expectation value $\left\langle \left[\frac{f(\mathbf{x})}{w(\mathbf{x})}\right] \right\rangle$ can be derived by using a set of N independent numbers which are distributed according to $w(\mathbf{x})$ and calculating

$$S = \frac{vol(B)}{N} \sum_{i=1}^{N} \frac{f(\mathbf{x})}{w(\mathbf{x})} \tag{3.75}$$

As stated earlier in this chapter one can derive for standard Monte Carlo integration the error from the Central Limit Theorem which is in this case:

$$\sigma_w(S(f/w)) = \sigma_w(f/w)\sqrt{N} \tag{3.76}$$

Therefore importance sampling is only helpful if $\sigma_w(f/w)^2 < \sigma(f)^2$ and so it is advisable to choose w in a way that the ratio f/w is approximately constant. The basic principle of this is illustrated in Figure 3.10. The distribution of the points in the integration phase space is changed according to the *importance* or magnitude of the integration function. As one is not able to just change the function to reduce the variance one can change the point density in a way that the set of points *sees* a flatter function. In practice the choice of the function w comes naturally by doing a variable transformation as will be shown in the next Section. In this work the problem is of the following form:

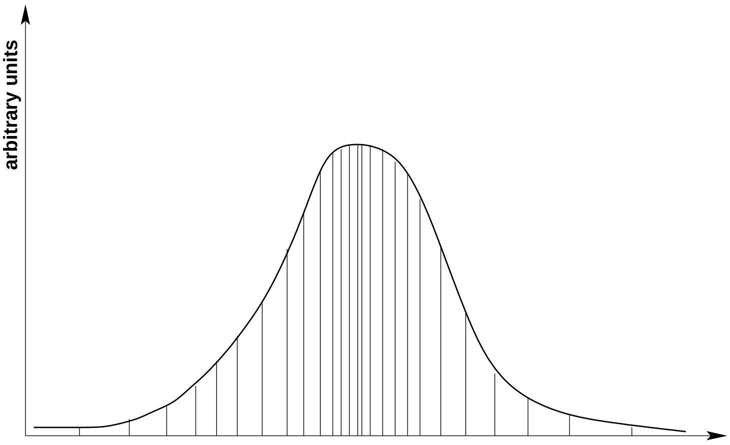

Figure 3.10: *The basic principle of the Importance Sampling method. The points used for integration are not distributed uniformly but follow the magnitude (importance of the integrand.*

$$I = P(l', p') = \int\limits_0^{l_{max}} \int\limits_0^{p_{max}} f(l,p) R_l(l'-l) R_p(p'-p)\, dl dp \tag{3.77}$$

which is a convolution in two dimensions of the physics function $f(l,p)$ using the resolution functions R_l and R_p which have a peak at 0 and can be characterized by their widths σ_l and σ_p which are small compared to the whole phase space $[0..l_{max}, 0..p_{max}]$.

Variable Transformation

The transformation is applied in two steps. The first is a shift to the mean value of the resolution l' and p' and an additional normalization to the expected widths σ_l and σ_p. The transformed variables over which to integrate are equally distributed in a volume B. The distribution is done by a grid in one dimension and a Niederreiter generator in the other. To map this region B onto the region $[0..l_{max}, 0..p_{max}]$ the transformation function was chosen to be the arctan function scaled appropriately, so that the variable transformation in total becomes:

$$x = \frac{1}{\pi} \arctan\left(\frac{l - l'}{\sigma_l}\right) + \frac{1}{2} \quad , \quad y = \frac{1}{\pi} \arctan\left(\frac{p - p'}{\sigma_p}\right) + \frac{1}{2} \tag{3.78}$$

and the original variables are:

$$l = l' + \sigma \tan(\pi(x - \frac{1}{2})) \quad , \quad p = p' + \sigma \tan(\pi(y - \frac{1}{2})) \tag{3.79}$$

Inserting this the original integral becomes:

$$I = \int\limits_{x_{min}}^{x_{max}} \int\limits_{y_{min}}^{y_{max}} f(l' + \sigma \tan(\pi(x - \frac{1}{2})), p' + \sigma \tan(\pi(y - \frac{1}{2}))) \cdot \tag{3.80}$$

$$R_l(\sigma \tan(\pi(x - \frac{1}{2})))R_p(\sigma \tan(\pi(y - \frac{1}{2}))) \cdot \tag{3.81}$$

$$\sigma_l \sigma_p \pi^2 \cos(\pi(x - \frac{1}{2}))^2 \cos(\pi(y - \frac{1}{2}))^2 \, dxdy \tag{3.82}$$

with transformed integration boundaries,

$$x_{min} = \frac{1}{\pi} \arctan\left(\frac{-l'}{\sigma_l}\right) + \frac{1}{2} \quad , \quad x_{max} = \frac{1}{\pi} \arctan\left(\frac{l_{max} - l'}{\sigma_l}\right) + \frac{1}{2} \tag{3.83}$$

$$y_{min} = \frac{1}{\pi} \arctan\left(\frac{-p'}{\sigma_p}\right) + \frac{1}{2} \quad , \quad y_{max} = \frac{1}{\pi} \arctan\left(\frac{p_{max} - p'}{\sigma_p}\right) + \frac{1}{2} \tag{3.84}$$

This transformation and its principle is illustrated in Figure 3.11. It shows for the decay length the equally distributed points in y direction and the points in the decay length. It is clearly visible that the density increases around the mean value $l' = 1$ cm in an area of $\pm\sigma_l = \pm 0.4$ cm. In Figure 3.12 the transformation principle is shown using the point sets from Figure 3.9. One can clearly see that the point density increases at the expected mean point of $[l', p'] = [1 \text{ cm}, 30 \text{ GeV}]$ with expected widths of $[0.4 \text{ cm}, 4 \text{ GeV}]$ for this example. The integration algorithm was tested in Karlsruhe with different point sets and the results can be found in [92].

3.3 Parameter Estimation

Many physics measurement problems are related in one form or another to the problem of estimating parameters from noisy data. A parameter estimation problem is usually formulated as an optimization one. Because of different optimization criteria and because of several possible parameterizations, a given problem

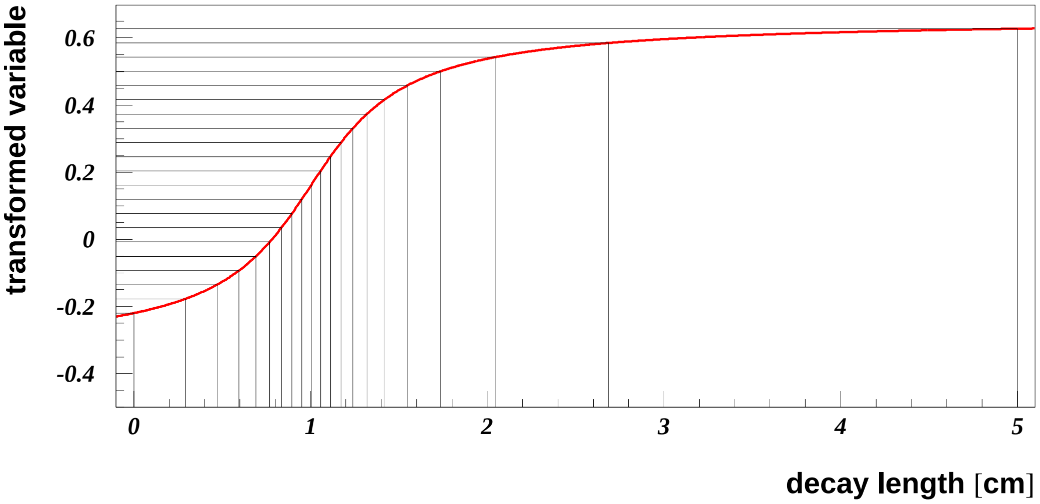

Figure 3.11: *The plot illustrates the principle of the variable transformation function. In this case a measured decay length of* $l' = 1$ *cm with a width of* $\sigma_l = 0.4$ *cm (to be able to see the effect) was considered. On the* y *axis the transformed variable is shown which is equally distributed (either by a grid or by quasi-random numbers). This leads to an increased point density around the point where the resolution peaks on the* x *axis.*

can be solved in many ways. The way chosen will influence the accuracy of the estimated parameters, the efficiency of computation, the robustness to predictable or unpredictable errors. This Section is focused on parameter estimation technique used in the analysis, namely the *method of least mean squares* and the *maximum likelihood method.*

3.3.1 Method of Least Mean Squares

The method of least mean squares has been developed in the 19th century and is a easy to handle and very powerful tool to analyze data with measurement errors. Due to these measurement errors the measured data values y_i differ from the *true* data values by an observable which is the standard deviation σ or variance σ^2. Assuming that there exists a functional relationship, a so-called *model* for the *true* data values, this model can depend on one or several parameters a_j for which no direct measurement exists. The method of least mean squares tries to minimize the square of the deviations of the measured values y_i from the model with respect to the parameters a_j.

$$S(a) = \sum_{i=1}^{n} \frac{(y_i - f_i(a))^2}{\sigma_i^2} = \text{minimum} \qquad (3.85)$$

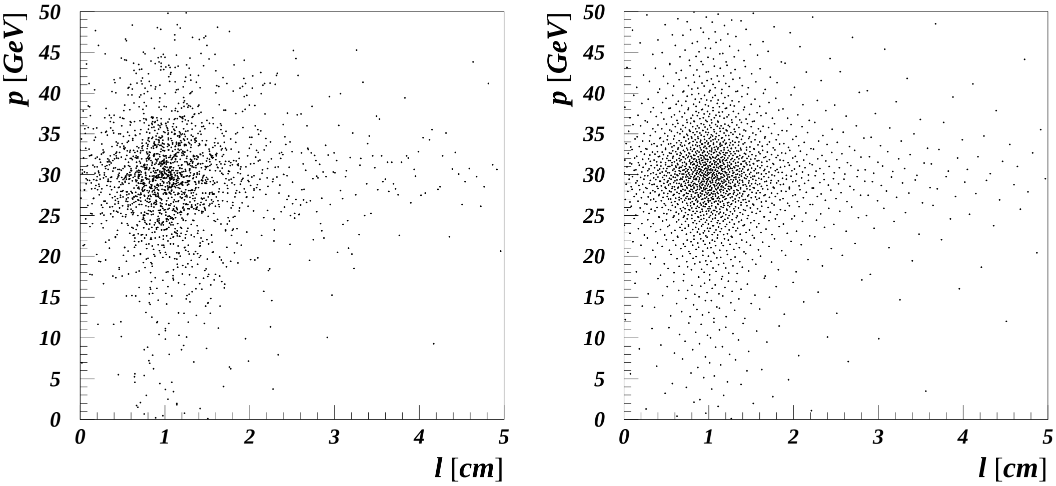

Figure 3.12: *These plots illustrate the effect of the variable transformation on the point sets from Figure 3.9 for mean values of* $l' = 1$ *cm and* $p' = 30$ *GeV and the expected widths* $\sigma_l = 0.4$ *cm and* $\sigma_p = 4$ *GeV respectively. Here also the numbers are chosen to make the effect more visible. In real life the widths are smaller and most of the mean decay length values are near the lower boundary.*

Here the f_i denote the model for the various y_i, from which also the variances σ_i^2 are known.

3.3.2 The Maximum Likelihood Method

The idea behind maximum likelihood parameter estimation is to determine the parameters that maximize the probability (likelihood) of the sample data. Although the methodology which will be described next for maximum likelihood estimation is simple, the implementation is mathematically delicate.

Assuming n measurements of the random variable x where x can be one single variable or a vector of variables with known probability density function $f(x|a)$, where a denotes one or more unknown parameters, the objective is to extract an estimate $\hat{a}$ of the parameters from the measurements of x.

The maximum likelihood method calculates the *Likelihood Function*

$$L(a) = f(x_1|a) \cdot f(x_2|a) \cdots f(x_n|a) = \prod_{i=1}^{n} f(x_i|n) \tag{3.86}$$

The function $L(a)$ is for a given set of measurements x_i a function of the parameter a and is an estimate of the probability to find these measured values for a certain parameter a. Therefore the best estimate of the parameter a is the value $\hat{a}$ for which the likelihood function is maximized, i.e. for which the probability is maximized to observe the values x_i. It is very important to ensure that the probability density function $f(x|a)$ is normalized for all possible values of a.

$$\int\limits_{\Omega} f(\mathbf{x}|a) = 1 \quad \forall\ a \tag{3.87}$$

Practically more convenient is the log of the likelihood function $F(a) = \ln L(a)$, which is called *Log-Likelihood Function*. The log-likelihood function is maximized for the same value $\hat{a}$ because the logarithmic function is monotonic. Usually one considers the negative log-likelihood function which then, of course, has to be minimized with respect of the parameters a. This can be done by common minimization algorithms like the Newton algorithm, gradient descent, line search or combinations of these. The error on the parameters $\mathbf{a}$ can be calculated as follows. Expanding the log-likelihood function around the minimum $\hat{\mathbf{a}}$ leads to [93]

$$\begin{aligned} F(\mathbf{a}) &= F(\hat{\mathbf{a}}) + \frac{1}{2}\sum_{i,j} \left.\frac{\partial^2 F}{\partial a_i \partial a_j}\right|_{\mathbf{a}=\hat{\mathbf{a}}} (a_i - \hat{a}_i)(a_j - \hat{a}_j) + \cdots \\ &= F(\hat{\mathbf{a}}) + \frac{1}{2}(\mathbf{a}-\hat{\mathbf{a}})^T H (\mathbf{a}-\hat{\mathbf{a}}) + \cdots \end{aligned} \tag{3.88}$$

or

$$\begin{aligned} L(\mathbf{a}) &\cong \mathrm{const} \cdot \exp\left(-\frac{1}{2}(\mathbf{a}-\hat{\mathbf{a}})^T H (\mathbf{a}-\hat{\mathbf{a}})\right) \\ &= \mathrm{const} \cdot \exp\left(-\frac{1}{2}(\mathbf{a}-\hat{\mathbf{a}})^T V^{-1} (\mathbf{a}-\hat{\mathbf{a}})\right) \end{aligned} \tag{3.89}$$

where V denotes the variance matrix and H the Hessian matrix of F. Therefore the variance matrix is the inverse of the Hessian.

The sum of squares of the method of least mean squares $S(a)$ and the negative log-likelihood function $F(a)$ are related by:

$$F(a) = \mathrm{const} + \frac{1}{2}S(a) \tag{3.90}$$

if the data are Gaussian distributed. From this relation follows that the variance of a is twice the inverse of the Hessian matrix of S. The practical problems of the maximum likelihood method arise from the fact that one has to know the model behind the measured values exactly. Another drawback with respect to computational minimization algorithms is the requirement of Equation 3.87 that for every iteration during the minimization process the normalization has to be calculated which usually can only be done numerically.

Chapter 4

Inclusive Reconstruction and Identification of b-Hadrons

The following chapter provides an overview of b-hadron reconstruction. For this purpose a reconstruction package called 'BSAURUS' [94] has been developed in Karlsruhe over the last 5 years. In this chapter only the parts of the BSAURUS package are described which were important for this analysis. BSAURUS makes extensive use of neural network techniques described in the previous chapter and is optimized for b hadron reconstruction while the enrichment of $b\bar{b}$ events was part of the official DELPHI b-tagging package (AABTAG) [95, 96].

One possibility to distinguish Z^0 decays into b quarks from decays into light quarks[1] is the identification of a high energetic muon or electron with high transverse momentum relative to the jet axis from the semileptonic decay of a b-hadron. However this method is limited by the semileptonic branching ratio of approximately 10% each. The usage of high resolution vertex detectors allows a more efficient tagging of $b\bar{b}$ events.
Due to the relative long lifetime of a b-hadron of $<\tau>\approx 1.55$ ps accompanied by a high boost due to the hard fragmentation ($<E_B>\approx 0.7\cdot E_{beam}\approx 30-35$ GeV), a b-hadron can fly several millimeters ($<l>\approx 3$ mm) before its decay. Figure 4.1 shows a $b\bar{b}$ event together with the vertex detector with two secondary vertices clearly displaced from the primary interaction point. The b-tagging makes use of this by defining a so-called 'impact parameter' for every track, which denotes the distance of the primary interaction point to the point of closest approach of the track. According to the projection of the impact parameter to the axis of the jet the track is assigned to, the impact parameter gets a sign (negative(positive) if the projection is 'behind'('in front of') the primary interaction point). As one can see from Figure 4.2 the impact parameter significance distribution of uds quarks is centered around zero, while it is shifted towards positive values for b

---

[1] light quarks denote u, d, s and c quarks in this context

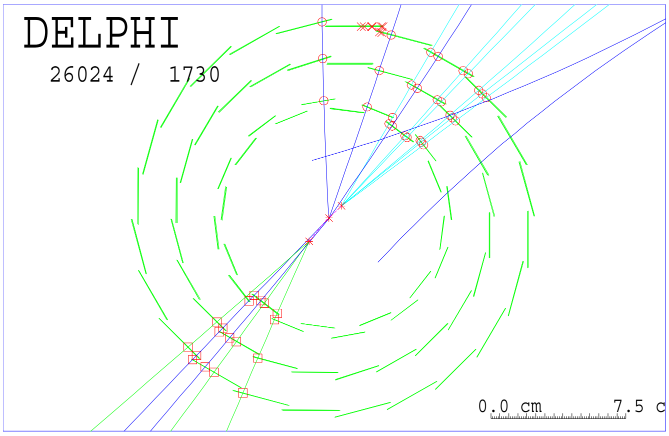

Figure 4.1: *A $Z^0 \to b\bar{b}$ event inside the* DELPHI *vertex detector. Both decay vertices are significantly separated from the primary vertex in the middle.*

quark events (c quark events are also slightly shifted due to the finite lifetime of c-hadrons). The impact parameter is then combined with other additional event information, e.g. vertex mass, rapidity, jet energy to get a good estimator for b quark identification. The first part of this chapter gives a more detailed description on b-tagging.

Essential for this analysis is a reconstruction of b events. The BSAURUS package provides estimators for the decay length and the momentum of the weakly decaying b-hadron. From this information the decay time can be calculated. Additionally the different b-hadron species have to be separated from the others and enriched to measure the individual lifetimes of b-hadrons.

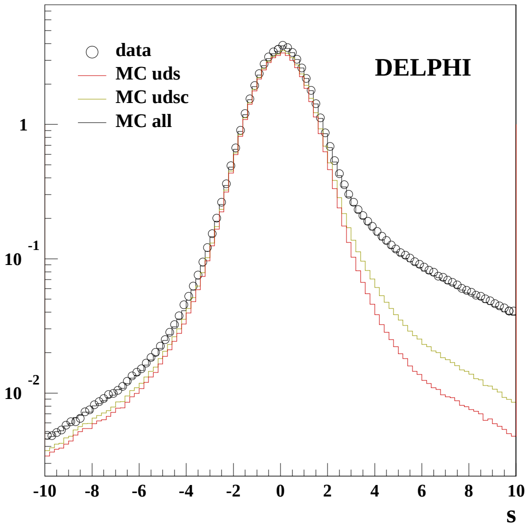

Figure 4.2: *The significance distribution of the impact parameter. Besides the data and Monte Carlo distributions the contributions of the different quark types are also shown. The surplus at positive values is used to separate $b\bar{b}$ events.*

The BSAURUS package starts with the selection of multihadronic events for which at least 5 charged particles and some energy deposit in the detector is required. An event is split into two hemispheres, following their two main jets. A first estimator for the b-hadron energy is provided by the 'rapidity algorithm'.

In addition the tracks selected by the rapidity algorithm are used to reconstruct the secondary vertex. Based on these quantities better estimators for the energy and decay length of the b-hadrons are calculated using neural network techniques. The second part of this chapter describes the different tools in detail. This part also contains the description of how different b-hadron species are separated from each other using a neural network.

4.1 b-Tagging

Using the impact parameter method to tag $b\bar{b}$ events requires a successful and precise reconstruction of the primary interaction point. The size and the position of the electron positron packages are used as a boundary condition. The position of the mean interaction point is calculated with 200 subsequent multihadronic events which allows a sufficient statistical precision and the consideration of possible movements of the particle packages during a fill [97]. The position can be estimated with a precision of 10 μm in x and 5 μm in y direction. The width in x direction depends on the run conditions and has typical values between 100 and 120 μm with an error of approximately 7 μm. The width in y direction is much smaller ($\leq$ 5 μm), because synchrotron radiation does not enlarge the width in this direction. To ensure that the reconstruction is not influenced by tracks from decay products of long living particles (K_S^0, Λ_0) or by interactions with the detector material, only tracks with more than one VD hit are used. These tracks are then used to minimize the following χ^2 function:

$$\chi^2(V_i) = \sum_a \frac{d_a^2}{\sigma_a^2} + \sum_{i=x,y,(z)} \frac{(b_i - V_i)^2}{(\sigma_i^b)^2}, \tag{4.1}$$

where d_a denotes the distance of the track to the primary vertex, σ_a the according error, V_i the coordinate i of the vertex, b_i the position of the mean interaction point and σ_i^b its width. In the years with Rz read-out of the VD (since 1994) the primary vertex is determined in three dimensions, before its position was only determined in $R\phi$. The fit is done iteratively, dropping the track with the biggest χ^2 contribution as long as the χ^2 value is above a certain threshold. The resolution in x and z direction (with Rz read-out of the VD) is for light quark events 22 μm and for $b\bar{b}$ events about 35 μm. The worse resolution in $b\bar{b}$ events is due to the fact that tracks from secondary vertices cannot completely be eliminated and that the multiplicity at the primary vertex is smaller for $b\bar{b}$ events.

With the help of this primary vertex one can calculate the impact parameter significance S for each individual track. S is defined as closest approach of the track to the primary vertex in the $R\phi$ plane divided by its uncertainty because the resolution of the VD in $R\phi$ is much better than in Rz. The significance S is signed positive (negative) if the angle between the vector from the primary vertex to the point of closest approach and the jet direction of the associated track is smaller (larger) than 90°. This leads to positive impact parameters for decay products of long lived particles. The impact parameters of tracks from the primary vertex are zero by definition. Due to limited detector resolution they should have positive and negative significances with equal probability. Figure 4.2 shows the significance distribution for data and Monte Carlo simulation and the contributions of the different quark types. The distribution for $S < 0$ is fixed by the detector resolution and has a Gaussian slope. Deviations from a Gaussian are caused by wrongly reconstructed tracks (e.g. wrong attachment of detector hits) and tracks from secondary decays and interactions. The negative tail of the significance distribution is used to determine the probability function $P(S_0)$ which denotes the probability that a track from the primary vertex has an absolute significance value bigger than S_0:

$$P(S_0) = \left\{ \begin{array}{c} \int_{S<S_0} f(S)dS, \text{ for } S_0 < 0 \\ P(-S_0), \text{ for } S_0 > 0. \end{array} \right. \tag{4.2}$$

where $f(S)$ is the resolution function determined by the significance distribution from data.With this track probability $P(S_0)$ one can calculate for a group of N tracks that all N tracks with their measured significances S_i stem from the primary vertex:

$$P_N = \Pi \cdot \sum_{j=0}^{N-1} (-\ln \Pi)^j / j! \quad , \text{ with } \Pi = \prod_{i=1}^{N} P(S_i). \tag{4.3}$$

The N-track probability P_N can be derived for different sets of tracks, e.g. for all tracks of an event or only the tracks of one hemisphere. P_N is smaller for events with a long lived particle (e.g. b quark and c quark decays), compared to uds production and gives therefore the possibility to enhance $b\bar{b}$ events by cutting on it.

4.1.1 The Confidence Method

The method described only uses the N-track probability to tag b events. By adding other variables the performance of the b-tagging can be enhanced. In the so-called 'Confidence Method' [98] besides the impact parameter significance the momenta of the tracks and the angles between the tracks and the jet direction are used. Their distributions for the different quark types are taken from Monte Carlo simulation. In contrast to the probability method of the previous section the confidence method tests the hypothesis that tracks have their origin in a certain quark type (uds, c or b). To enhance b events values around one for the b hypothesis are required.

4.1.2 The Combination Method

The b-tagging algorithm can be further improved by making use of the expected jet structure of the b quark and additional discriminating variables [99]. Therefore the jet structure of the event was resolved using the JADE algorithm [56] with jet resolution parameter $y_{cut} = 0.01$. The combination method combines different, discriminating variables x_i to one single variable y. Considering independent variables one can determine y in the following way:

$$y = \frac{f^B(x_1, \ldots, x_n)}{f^S(x_1, \ldots, x_n)} = \prod_{i=1}^{n} \frac{f^B(x_i)}{f^S(x_i)} = \prod_{i=1}^{n} y_i \quad , \tag{4.4}$$

where $f^B(x_i)$ and $f^S(x_i)$ denote the probability density functions for background and signal. The procedure is the following:

- for each discriminant variable the probability density functions for signal and background ($f^B(x_i)$ and $f^S(x_i)$) are determined
- new variables $y_i = \frac{f^B(x_i)}{f^S(x_i)}$ are calculated
- the combined variable $y = \prod_{i=1}^{n} y_i$ is calculated
- events with $y < y_0$ are tagged as signal, where y_0 can be varied to select desired purities or efficiencies of the tagging.

This approach is also used in the confidence method. In real life the variables are not independent, so that the probability density functions depend on all n variables, which makes it difficult to determine the functions for $n > 2$. In the combination method variables are chosen, which are as independent as possible. Used variables are:

- The most important variable is the N-track probability P_N (P_j^+ in the plot) based on the impact parameter significance described in the previous section.
- The sum of four-momentum-vectors of all tracks associated to a jet is calculated. The invariant mass M_s is derived from the jet four-momentum. For c quarks a sharp drop at $M_s = 1.8$ GeV/c^2 is expected, while the distribution extends up to $M_s = 5$ GeV/c^2 for b quarks.
- The rapidity $R_s^{tr} = \frac{1}{2}\ln(\frac{E+p_L}{E-p_L})$ is expected to be bigger for c quarks compared to b quarks due to the lower mass and the lower multiplicity.
- The energy fraction X_s^{ch} of charged particles from the secondary vertex compared to all tracks in the jet is usually higher for b quarks compared to light quarks.

The two backgrounds, jets from c and from light quarks, are independent and different from each other. Therefore the combined b-tag variable can be written as:

$$y = n_c \cdot \prod \frac{f^c(x_i)}{f^b(x_i)} + n_q \cdot \prod \frac{f^q(x_i)}{f^b(x_i)} \quad = \quad n_c \cdot \prod y_i^c + n_q \cdot \prod y_i^q \tag{4.5}$$

where n_c and n_q denote the fraction of c and uds jets with reconstructed secondary vertex and $f^q(x_i)$, $f^c(x_i)$ and $f^b(x_i)$ denote the probability density functions of the variables x_i for uds, c and b quark jets respectively [100]. Figure 4.3 shows the efficiency vs. purity curve for the individual contributions of the discriminant variables. One can see that especially the inclusion of the rapidity and the invariant mass improves the selection at very high purities.

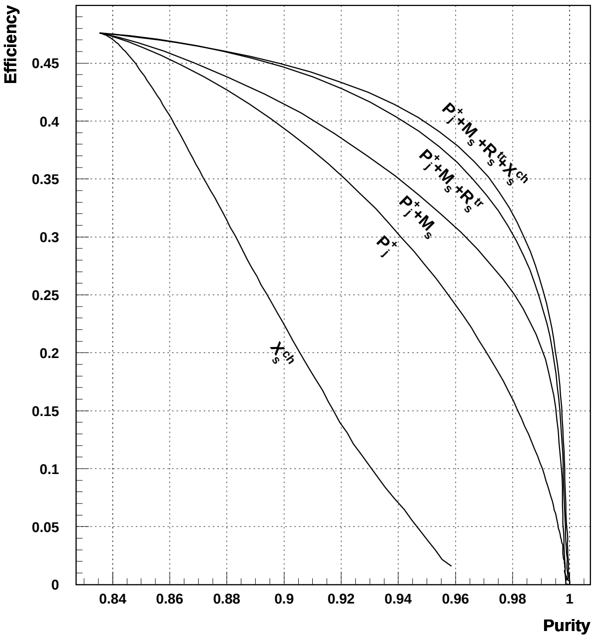

Figure 4.3: *The plot illustrates the efficiency vs purity behavior of the combined* DELPHI *b-tagging calculated for quark jets. By adding extra information from invariant mass M_s, rapidity R_s^{tr} and charged energy fraction X_s^{ch} to the pure impact parameter tag, the performance is improved dramatically.*

4.2 BSAURUS

One of the properties of b-hadrons is their large number of decay channels (several thousands) each with a small branching ratio. Standard analyses select specific decay channels with a clear signature. Using these exclusive channels results, however, in limited statistics and hence to a severe limitation for studying b physics. The idea of BSAURUS is to reconstruct inclusively, which means regardless of decay channels, as many properties of b jets as possible with high efficiency and good purity. This is achieved by using physics knowledge about b-hadron production and decays wherever possible and by exploiting the capabilities of the DELPHI detector to their extreme. To get the maximum out of the different sources of information they are combined using neural network techniques.

The following section gives a step by step overview of the main processing stages of BSAURUS, where the focus is on properties which are essential for this analysis.

4.2.1 Multihadronic Selection

BSAURUS starts with an event selection loose enough to provide sufficient freedom for individual analyses to apply their selection. However, to ensure that all quantities can be calculated successfully one has to apply a minimal hadronic selection. The aim of this selection is the effective suppression of the following background processes:

- leptonic Z^0 decays, in particular $\tau^+\tau^-$ events which may have a multi hadron topology as well
- photon-photon events
- interactions of the beam with rest gas or the beam pipe

Multihadronic Z^0 events were selected using the following requirements:

- at least 5 charged particles with momenta larger than 400 GeV/c, a polar angle between 20° and 60° and a track length of at least 30 cm have to be reconstructed
- at least 3% of the total energy of the charged particles has to be in each of the forward and backward hemispheres defined with respect to the beam axis, where only tracks with a momentum larger than 0.2 GeV/c are counted
- the total charged energy sum of the whole event should be at least 12% of the center of mass energy, where pion masses are assumed for all particles to ensure a conservative hypothesis

These requirements, taken from the DELPHI pilot record hadron selection, ensure a selection of multihadronic events with an efficiency of more than 95% with a remaining background of less than 1%. The same multihadronic selection criteria are required for the Monte Carlo sample, based on fully simulated $Z^0 \rightarrow q\bar{q}$ events using JETSET 7.3 [56].

4.2.2 Standard Particle Selection

Charged particles used by the BSAURUS algorithms must pass the following track selection criteria:

- impact parameter, with respect to the origin in the $R\phi$ plane
  $|\delta_{R\phi}| < 4$ cm
- impact parameter, with respect to the origin in the z plane
  $|\delta_z| < 6$ cm
- $|\cos\theta| < 0.94$
- $\frac{\Delta E}{E} < 1.0$
- at least one $R\phi$ track from the vertex detector (VD)
- tracks must not have been flagged as originating from interactions with material

Neutral particles are included that are flagged by the DELPHI mass code as being a photon, π^0, K_S^0 or Λ^0

4.2.3 Event Jets

Event jets are reconstructed via the routine LUCLUS [101]. The distance of two clusters[2] is defined as:

$$d_{ij}^2 = \frac{4 \cdot |\vec{p}_i|^2 \cdot |\vec{p}_j|^2 \cdot \sin^2(\theta_{ij}/2)}{(|\vec{p}_i| + |\vec{p}_j|)^2} \tag{4.6}$$

For $d_{ij} < d_{join}$ two clusters are combined. At LEP energies a transverse momentum cutoff value of $d_{join} = 5.0$ GeV/c is used. To define the orientation of an event, the property thrust is used:

$$T = \max_{|\vec{n}|=1} \frac{\sum_i |\vec{n} \cdot \vec{p}_i|}{\sum_i |\vec{p}_i|}. \tag{4.7}$$

The thrust axis is defined by the unity vector $\vec{n}$ for which T becomes maximal. Values of T lie between 1/2 for isotropic (spheric) events and 1 for two-jet events.

---

[2] a cluster consists of one or more particles

4.2.4 Event Hemispheres

All reconstructed particles, that pass the standard selection, are used to calculate the event thrust axis described in the previous section via routine LUTHRU [101]. Event hemispheres are then defined by the plane perpendicular to the thrust axis.

Each hemisphere has associated with it two axes: the thrust axis and the reference axis used for the calculation of rapidity (see Section 4.2.5 below) for tracks in that hemisphere. The reference axis is a jet axis in the following way:

- In the case of a two jet event, the reference axis for the hemisphere is the jet axis in that hemisphere.
- If a hemisphere contains 2 or more jets (i.e. an event with 3 or more jets):
  - If one of the jets is the highest energy jet in the event, that jet axis forms the reference axis.
  - If the highest energy jet is in the opposite hemisphere, the combined probability (P_{jet}) for the tracks in a jet to have originated from the event primary vertex is formed (via the AABTAG package algorithms [95, 96, 99]).
    The jet that is most 'B-like', i.e. with the smallest probability, is then selected if $P_{jet} < 0.5$.
  - If no jet in the hemisphere satisfies the above criteria, the jet with the highest energy is selected to form the reference axis.

In addition, internally to BSAURUS, the hemispheres are numbered as 1 or 2 with the convention that the jet forming the reference axis for hemisphere 1 is of higher energy than that for hemisphere 2.

4.2.5 *B* Energy Reconstruction

The Rapidity Algorithm

The rapidity of a track is defined as follows,

$$y = \frac{1}{2} \ln(\frac{E + p_L}{E - p_L}) \tag{4.8}$$

for E the track energy and p_L the longitudinal momentum component along the reference axis for the hemisphere. The rapidity has the property that a Lorentz transformation modifies it by just adding a constant, i.e. $y' = y + f(\beta)$. Particles from the b decay are peaked at a higher rapidity than fragmentation particles (see Figure 4.4). The so-called 'rapidity algorithm' returns an initial estimate of the

b-hadron candidate four-vector. This is defined simply to be the sum of individual track momentum vectors in a hemisphere for tracks with rapidity greater than 1.6. The value of the cut provides a good rejection of tracks originating from the primary vertex while accepting the majority of tracks from the b decay.

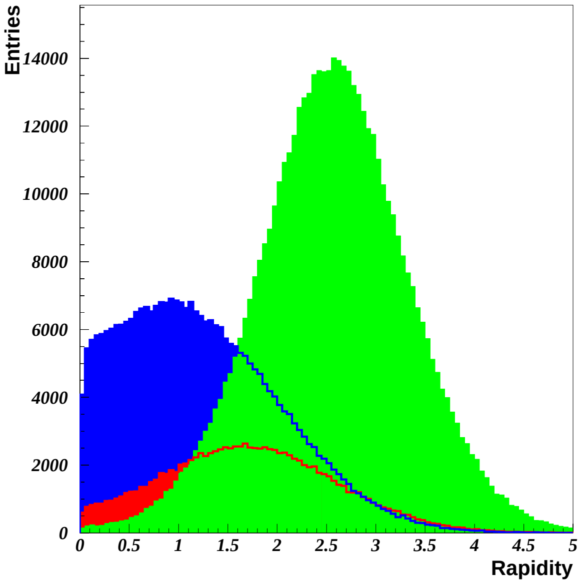

Figure 4.4: *The b-hadron rapidity distribution. Tracks are selected from events that are contained in the barrel and with 2-jets at a* $Z^0 \rightarrow b\bar{b}$ *purity of about* 80%. *The three different shaded areas indicate particles originating from a weakly decaying b-hadron (light grey), fragmentation partner of the b-hadron (dark grey) and from the fragmentation (black).*

Energy Correction

There are still badly reconstructed events which can be rejected by applying the following requirements:

- the initial reconstructed b candidate energy E_{raw} is 20 GeV or more
- the initial reconstructed b candidate mass m_{raw} lies within two standard deviations of the total sample median value
- the ratio, x_h, of the hemisphere energy E_{hem} to beam energy E_{beam} lies in the range $0.6 < x_h < 1.1$

Applying these cuts leads to a reduction of b events of 26%. Monte Carlo studies show a strong correlation between the generated b energy E_{truth} and the reconstructed energy E_{raw}. Additional correlations exist between the energy residuals $\Delta E = E_{raw} - E_{truth}$ and m_{raw} (which is approximately linear in m_{raw}) and between ΔE and x_h caused by energy losses due to neutral particles and inefficiencies. To improve the energy resolution and to get a better estimator for the energy a correction is applied which accounts for these inefficiencies and neutral energy losses. The starting point of this correction are the initial estimates of the b energy and mass, E_{raw} and m_{raw}. Following Monte Carlo studies, these estimates are chosen to be from the rapidity algorithm for events with > 2 jets and where $x_h > 0.6$, and to be derived from the sum of 'b weighted' four-vectors otherwise. This involves weighting (via a sigmoid threshold function) the momentum and energy components of charged tracks by the weights derived from a neural network, called 'TrackNet', which separates fragmentation tracks from b-hadron decay tracks (see Section 4.2.8). Neutral particles are weighted by their rapidity. In this way the effect of tracks from the b decay are enhanced and tracks from the primary vertex are suppressed. The correction proceeds in the following way: The data are divided into several samples according to the measured ratio x_h and for each of these classes the b energy residual ΔE is plotted as function of m_{raw}. The median values of ΔE in each bin of m_{raw} are calculated and their m_{raw} dependence fitted by a third order polynomial

$$\begin{aligned} \delta(m_{raw}; x_h) \;&=\; a + b(m_{raw} - < m_{raw} >) + c(m_{raw} - < m_{raw} >)^2 \\ &\quad + d(m_{raw} - < m_{raw} >)^3 \end{aligned} \tag{4.9}$$

The four parameters a, b, c and d in each x_h class are then plotted as a function of x_h and their dependence fitted with third and second-order polynomials. Thus one obtains a smooth correction function describing the mean dependence on m_{raw} and the hemisphere energy as determined from the Monte Carlo. Finally, a small bias correction is applied for the remaining mean energy residual as a function of the corrected energy.

In addition, for the case of hemispheres with larger missing energy, i.e. $x_h < 0.6$, a further correction is derived separately for hemispheres 1 and 2.

The procedure described above leads to a single estimate of the b-hadron energy. The resolution of the reconstructed b-hadron energy is shown in Figure 4.5 together with the resolution of the raw energy estimator from the rapidity algorithm alone.

4.2.6 Inclusive Secondary Vertex Finding

To fit a secondary vertex tracks of the observed hemisphere are selected which pass the track selection criteria of Section 4.2.2 and with rapidity > 1.6. To this class of tracks, additional criteria are applied with the aim of selecting tracks for the vertex fitting stage that are likely to have originated from the decay chain of a weakly decaying b-hadron state. It is important to reject as far as possible tracks from the fragmentation in order to avoid large pulls in the vertex position toward the primary vertex. To apply a secondary vertex fit at least two tracks have to be selected. The selection process is as follows:

1) the highest energy muon or electron candidate is selected if $E_{lepton} > 3$ GeV

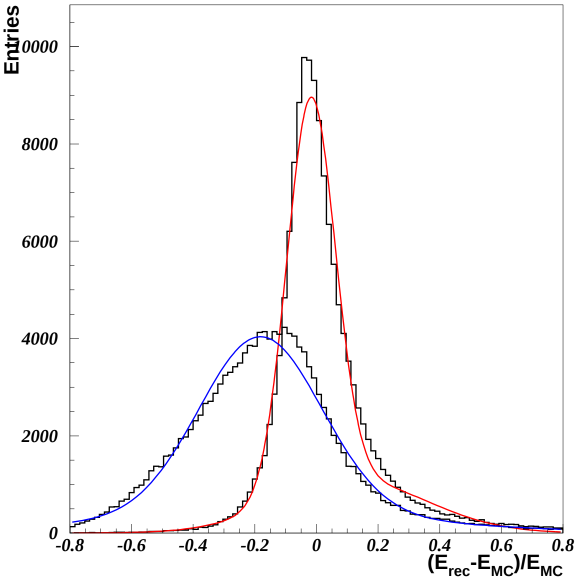

Figure 4.5: *The resolution of the reconstructed b-hadron energy for the raw energy estimator from the rapidity algorithm (black curve) and the corrected energy estimator (light gray). A double Gaussian is fitted to the distributions with widths* $\sigma_1 = 0.20$ *and* $\sigma_2 = 0.54$ *for the inner and outer Gaussian respectively for the raw estimator and* $\sigma_1 = 0.08$ *and* $\sigma_2 = 0.24$ *for the corrected one.*

2) the crossing point of each track with the initial b candidate direction vector (see Section 4.2.5) is found and the distance, L, between primary vertex[3] and crossing point is calculated. Tracks are selected if $L/\sigma_L > 2$ or, for cases where $L/\sigma_L < 2$, tracks must satisfy $y > 2.5$ and $L > 0.1$ cm as well.

If at this stage the number of tracks selected for the fit is less than 2, further attempts are made to add tracks to the fitting procedure. When this happens in $b\bar{b}$ events, it is likely that the true decay length of the b-hadron was small. If only one track was found in the previous stages:

3) add tracks with $y > 3$

4) if still only one track is selected, add the track of highest rapidity from the remaining track list of the hemisphere

If no track was selected from the previous stages:

5) A search is made for the best kaon candidate in the hemisphere based on a neural network (included in the MACRIB [102] package) using information from particle identification components of the detector. If this kaon candidate has $y > 2$ it is added to that track in the remaining track list of the hemisphere that has the highest rapidity. This track pair alone will then form the starting track list for the vertexing procedure

---

[3] Unless otherwise stated, the primary vertex definition in BSAURUS is the event primary vertex from the AABTAG package.

6) Finally if no kaon candidate exists in the hemisphere, the two tracks of highest rapidity are selected.

Using the track list supplied by this track selection process, a secondary vertex fit is performed in 3 dimensions (using routine DAPLCON from the ELEPHANT package [103]) constrained to the direction of the B candidate momentum vector. The event primary vertex is used as a starting point and if the fit is not acceptable[4], the track making the largest χ^2 contribution is stripped away in an iterative procedure and the fit repeated.

In addition to returning the secondary vertex position, the procedure also fits the primary vertex position and updates the B candidate direction according to the vector joining the primary and secondary vertex points. This information is used in forming some of the input of the TrackNet as described in Section 4.2.8.

Once a convergent fit has been attained, the final stage of the secondary vertex fitting procedure involves an attempt to add into the fit tracks that failed the initial track selection criteria but nevertheless are consistent with originating from the vertex. These tracks are identified on the basis of an 'intermediate' version of the TrackNet. This is a neural network output that discriminates between tracks originating from the primary vertex and those likely to have come from a secondary vertex. An intermediate version of the TrackNet is constructed, specifically for the purpose of use in this final stage of vertex fitting, based on secondary vertexing information available before this final stage has run. In general, secondary vertex tracks have TrackNet output $\simeq 1$ whereas tracks from the primary vertex have TrackNet values close to zero. The track of largest TrackNet output in the hemisphere is added to the existing track list and retained if the resulting fit converges. This process continues iteratively for all such tracks with TrackNet output > 0.5. The resolution of this standard BSAURUS vertex algorithm is shown in Figure 4.6.

4.2.7 Decay Length and Decay Time

The b-hadron decay length estimate in the $R\phi$ plane , $L_{R\phi}$, is defined as the (positive) distance between the primary and secondary vertex positions reconstructed from the secondary vertex search described in the previous Section. The 3-dimensional decay length is constructed as $L = L_{R\phi}/\sin(\theta)$, where the b-hadron candidate momentum vector, P_b detailed in Section 4.2.5, defines the direction. The resulting proper decay time estimate is given by $t = Lm_b/p_b c$, where the b rest mass is taken to be 5.2789 GeV/c^2.

---

[4] Here, 'not acceptable' means the fit took more than 20 iterations. A further iteration is deemed necessary if the χ^2 is above 4 standard deviations during the first 10 iterations or above 3 standard deviations during the next 10 iterations.

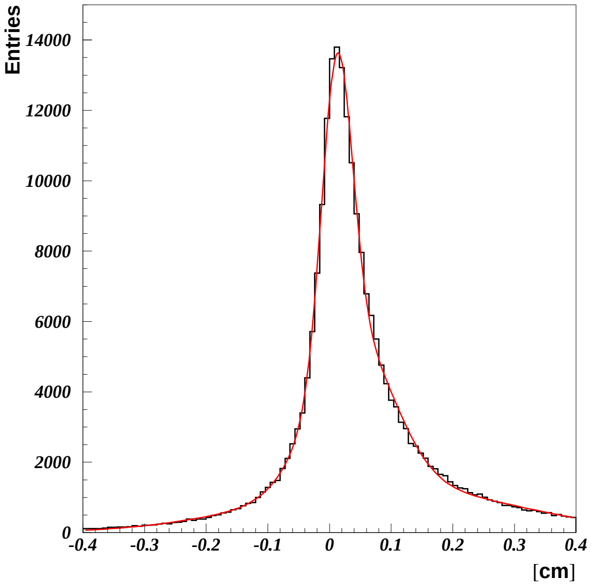

Figure 4.6: *The decay length resolution for the standard vertex algorithm of* BSAURUS *is illustrated. The rather large forward bias is clearly visible. Three Gaussians were fitted to the distribution with:*

$$\begin{aligned} \bar{l}_1 &= 122\,\mu m & \sigma_1 &= 246\,\mu m \\ \bar{l}_2 &= 385\,\mu m & \sigma_2 &= 648\,\mu m \\ \bar{l}_3 &= 941\,\mu m & \sigma_3 &= 2039\,\mu m \end{aligned}$$

4.2.8 The b-Track Probability Network (TrackNet)

The rapidity algorithm separates already tracks from fragmentation and from the b decay chain quite good. To get a better separation several discriminating variables are combined via a dedicated neural network, the 'TrackNet', which calculates for every track the probability of the track originating from a weak b decay.

Only tracks passing the standard quality cuts of Section 4.2.2 are considered. The discriminating variables that form the input to the network per track are:

- the track total momentum
- the track momentum in the b candidate rest frame
- the helicity angle of the track defined as the angle between the track vector in the b candidate rest frame and the b candidate momentum vector in the lab frame
- a flag to identify whether the track took part in the secondary vertex fit or not
- the probability that the track originates from the fitted primary vertex (AABTAG algorithm)
- the probability that the track originates from the secondary vertex (AABTAG algorithm)

- the probability that the track originates from the fitted primary vertex (Bsaurus algorithm)
- the probability that the track originates from the secondary vertex (Bsaurus algorithm)
- the track rapidity

In addition, input variables that gave no inherent discriminating power were included to inform the network of the potential quality of the other input variables:

- The decay length or distance between the primary and secondary vertex in the $R\phi$ plane. This quantity is then scaled by the reconstructed error to form a decay length significance.
- the track quality word which contains flags according to ambiguities, tracks coming from secondary interactions, etc. (for more details see [94])

For the training of the network tracks from fragmentation where attached to the network output zero and tracks from the b decay chain where attached to the network output 1. The network output of the TrackNet is illustrated in Figure 4.7. One can clearly see the good separation power between b decay and fragmentation tracks.

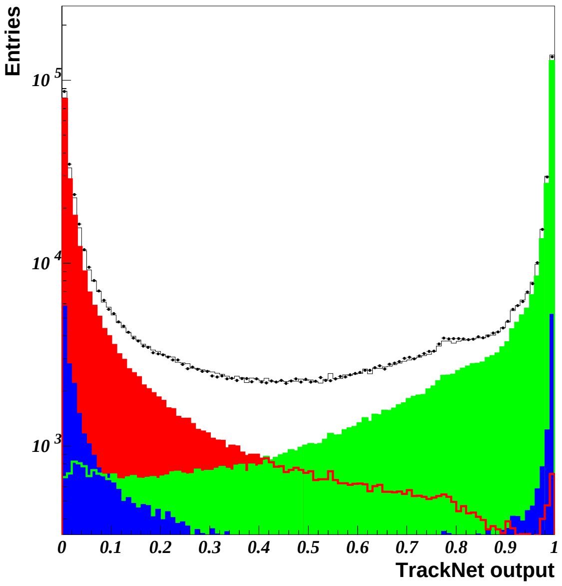

Figure 4.7: *The TrackNet output distribution for Monte Carlo compared to the data. Tracks are selected from 2-jet events that are contained in the barrel and are about 90% pure in $Z^0 \rightarrow b\bar{b}$ events. The three shaded areas indicate particles originating from the b-hadron decay chain (light grey), fragmentation or excited b-hadron decay (dark grey) and any tracks from non-b-decays of the Z^0 (black).*

4.2.9 Vertex Reconstruction

In Figure 4.6 the decay length resolution of the standard Bsaurus vertex algorithm of Section 4.2.6 is shown. Clearly visible is the rather large bias in the direction of larger decay lengths. Starting from the standard secondary vertex described in Section 4.2.6, four independent algorithms were implemented in Bsaurus with the aim of improving the decay length resolution and minimizing any forward bias resulting from the inclusion of tracks from the vertex of the cascade D decay in the b decay vertex reconstruction. These algorithms were based on a neural network (BDnet) designed to discriminate between tracks whose origin was the weakly decaying b-hadron and tracks that originated from the subsequent decay of the cascade D meson[5], This section gives a brief description of these four algorithms after explaining the B-D separation net (BDnet).

B-D Separation (BDnet)

In selecting tracks for inclusion in the b secondary vertex fit described in Section 4.2.6, there is inevitably some background from tracks that originate not from the b decay vertex directly, but from the subsequent D cascade decay point. The effect of including such tracks is illustrated in Figure 4.8 exhibiting that, on average, the b decay length estimate reconstructed by Bsaurus lies somewhere between the true b decay point and that of the D. The BDnet is a neural network designed to discriminate between tracks originating from weakly decaying b-hadrons and

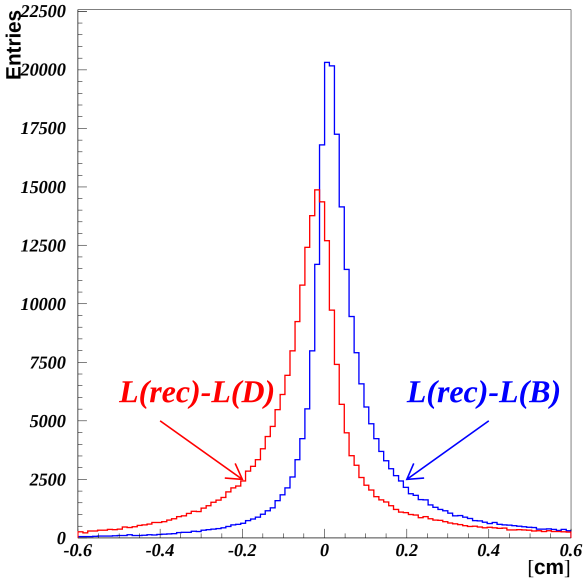

Figure 4.8: *Comparison of $(L(rec) - L(B))$ to $(L(rec) - L(D))$ illustrating that the reconstructed decay length is often larger than the true b decay length but less than the true D cascade decay length. The origin of this pull to larger decay lengths is the inclusion of one or more tracks from the D vertex.*

---

[5] A B meson usually decays to a D meson: $B \to D \to K$. This is called a decay cascade.

those from subsequent D meson decay cascade. The following discriminating variables are used:

- the angle between the track vector and the estimate of the flight direction of the b which was derived from the b-hadron four-momentum vector (see Section 4.2.5)
- the probability that the track originates from the fitted primary vertex (AABTAG algorithm)
- the probability that the track originates from the fitted secondary vertex (AABTAG algorithm)
- the momentum and angle of the track vector in the b rest frame
- the TrackNet output
- kaon identification information (KaonNet see [102] for details)
- lepton identification information

The remaining variables carry no implicit discriminating power but are included as gauges of the quality of the other variables

- the track quality word (already mentioned in Section 4.2.8)
- the hemisphere quality word for the whole hemisphere which contains flags according to e.g. the number of secondary interactions, tracks containing detector hits that could equally well fit to other tracks in the vertex detector (so-called ambiguities) or the number of tracks in the hemisphere which are excluded by selection etc. (for more details see [94])
- the decay length significance of the hemisphere, L/σ_L
- the secondary vertex mass of the hemisphere
- the hemisphere rapidity gap between the track of highest rapidity below a TrackNet cut at 0.5 and the smallest rapidity above the cut at 0.5.

The output of the BDNet is shown in Figure 4.9 for simulation and data for tracks with TrackNet > 0.5. The two classes the network was trained on are shown in the histogram, namely tracks originating from cascade D decays (light grey) and all other tracks which are mainly tracks from the b decay (dark grey). The main reason for the small discrepancy between data and simulation are the different D branching ratios which are not known with sufficient precision.

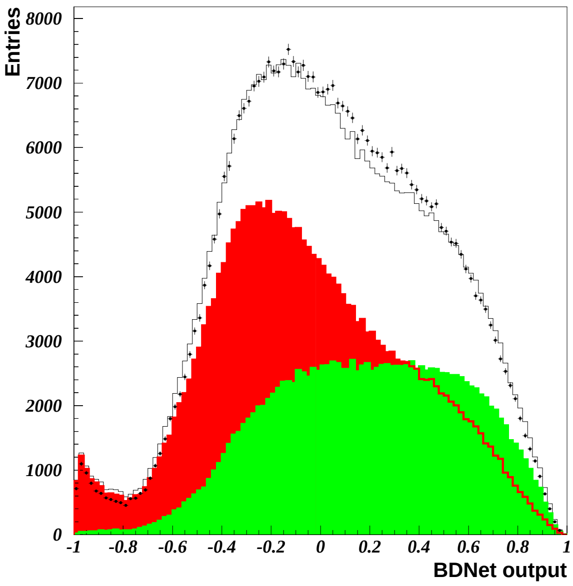

Figure 4.9: *Output of the BDNet for tracks in simulation (histogram) and data (points), for tracks with TrackNet* > 0.5. *The component histograms show the distribution for cascade D decay tracks (light grey) and B decay tracks (dark grey).*

The Strip-Down Method

The basic idea of the Strip-Down algorithm is to select candidate tracks most likely to come from the weakly decaying b, excluding tracks from the decay cascade. Therefore candidate tracks are selected if they have a TrackNet output larger than 0.5 and BDNet value less than zero. Monte Carlo studies have shown that this BDNet cut gives an optimal decay length resolution for the Strip-Down fit.

A secondary vertex fit is performed if 2 or more tracks are selected. If the fit is not acceptable (within the same criteria as for the standard BSAURUS secondary vertex fit described in Section 4.2.6) and more than two tracks are originally selected, the track with the highest χ^2 contribution is removed[6] and the fit repeated. This procedure continues iteratively until convergence is reached or only two tracks are left. The fit is constrained by the direction estimated from the b-hadron four-momentum reconstruction and the starting point of the fit are the secondary vertex coordinates of the standard fit.

Improving the resolution by this way, however, has a draw-back. In a lot of events less than two candidate tracks are selected or the fit with the two finally selected tracks did not converge. This leads to an overall efficiency for this algorithm of around 44% in b events depending slightly on event selection criteria.

---

[6] This stripping procedure of tracks from the fit is responsible for the name of the algorithm

The Build-Up Method

In the Build-Up algorithm a seed vertex was formed by those two tracks with TrackNet bigger than 0.5 and smallest BDNet values. The two selected tracks have the highest probability to come from a weakly decaying b-hadron. If the invariant mass of all remaining tracks with TrackNet > 0.5 exceeds the D mass[7], that track with the lowest BDNet output is added to the seed vertex definition and was also fitted to a common vertex with the two seed tracks. This process continued iteratively until either the fit failed to converge or the mass in remaining tracks dropped below the D meson mass. Finally all the candidate tracks are fitted to a common vertex.

The Build-Up vertex has an efficiency of about 83% in b events which is better than for the Strip-Down algorithm. However, it has the drawback that the bias from the D vertex is hardly removed.

The D-Rejection Method

In the D-rejection method, a vertex of the candidate cascade D was built by applying the opposite BDNet selection criteria to the Strip-Down method i.e. candidate tracks were selected if they satisfied TrackNet output bigger than 0.5 and BDNet value greater than zero. The fit started with two tracks. Then additional tracks were added until the fitted vertex exceeded the D mass. The b candidate vertex was then fitted using the Strip-Down algorithm but applied to all tracks *except* those already selected for the D vertex. Here the efficiency for this algorithm is around 40%.

The Semileptonic Algorithm

The semileptonic algorithm attempted to improve the vertex resolution for the case of semileptonic decays of b-hadrons where energy has been carried away by the associated neutrino. When there was a clear lepton candidate in the hemisphere, the algorithm reconstructed a cascade D candidate vertex in a similar way to the D-rejection method but with the lepton track excluded. The tracks associated with the vertex were then combined to form a 'D candidate track' which was extrapolated back to be vertexed with the lepton track and so make the b candidate vertex.

In only 16% of all b events this method converges due to the fact that only $\approx 21\%$ of all b-hadrons decay semileptonically.

---

[7] A D mass is taken to be 1.7 GeV/c^2.

Comparison of the Different Vertex Algorithms

Figure 4.10 illustrates the four different vertex algorithms in comparison with the standard BSAURUS algorithm. In each plot only b events where used and only the subset of events where the vertex algorithm converged. The distributions were then normalized according to the standard algorithm. One can clearly see, that the Strip-Down fit removes the forward bias. Also visible is a slight improvement of the resolution. Also for the other vertices the resolution is slightly improved.

As stated earlier the overall efficiency for the different vertices are 44%, 83%, 40% and 16% for the Strip-Down algorithm, the Build-Up Method, the D rejection method and the semileptonic algorithm respectively. The combined efficiency of requiring either one of the first three algorithms (the semileptonic algorithm is only accessible for semileptonic decays) is 83%, which shows that around 17% of the events are not accessible to either one of the algorithms. Monte Carlo studies have shown that those events often have badly reconstructed tracks and their resolution within the standard algorithm is bad. These events have a mean resolution of approximately 780 μm with a huge tail and are not usable in the analysis.

4.2.10 The b-hadron Identification Network

The following section describes an attempt to decide in each hemisphere whether the hemisphere contains a $B+$, B^0, B_s or a b-baryon[8].

For this purpose a neural network consisting of 15 input nodes, described in detail below, 17 nodes in the single hidden layer and 4 output nodes was trained. Each output node delivers a probability for the hypothesis it is trained on i.e. the first supplies the probability for a B_s meson to be produced in the hemisphere, the second for a B^0 meson, the third for charged B^+ mesons and the fourth for all species of b-baryons. One of the input variables is the secondary vertex charge which will now be described in more detail than the others.

The Inclusive Secondary Vertex Charge

The weighted vertex charge is formed using the TrackNet value as a probability P_B for each track of one hemisphere to originate from the b-hadron decay vertex rather than from the primary vertex:

$$Q_B = \sum_{i}^{N_{hem}} P_B(i) \cdot Q(i) \tag{4.10}$$

---

[8] The charge conjugate states are also implied

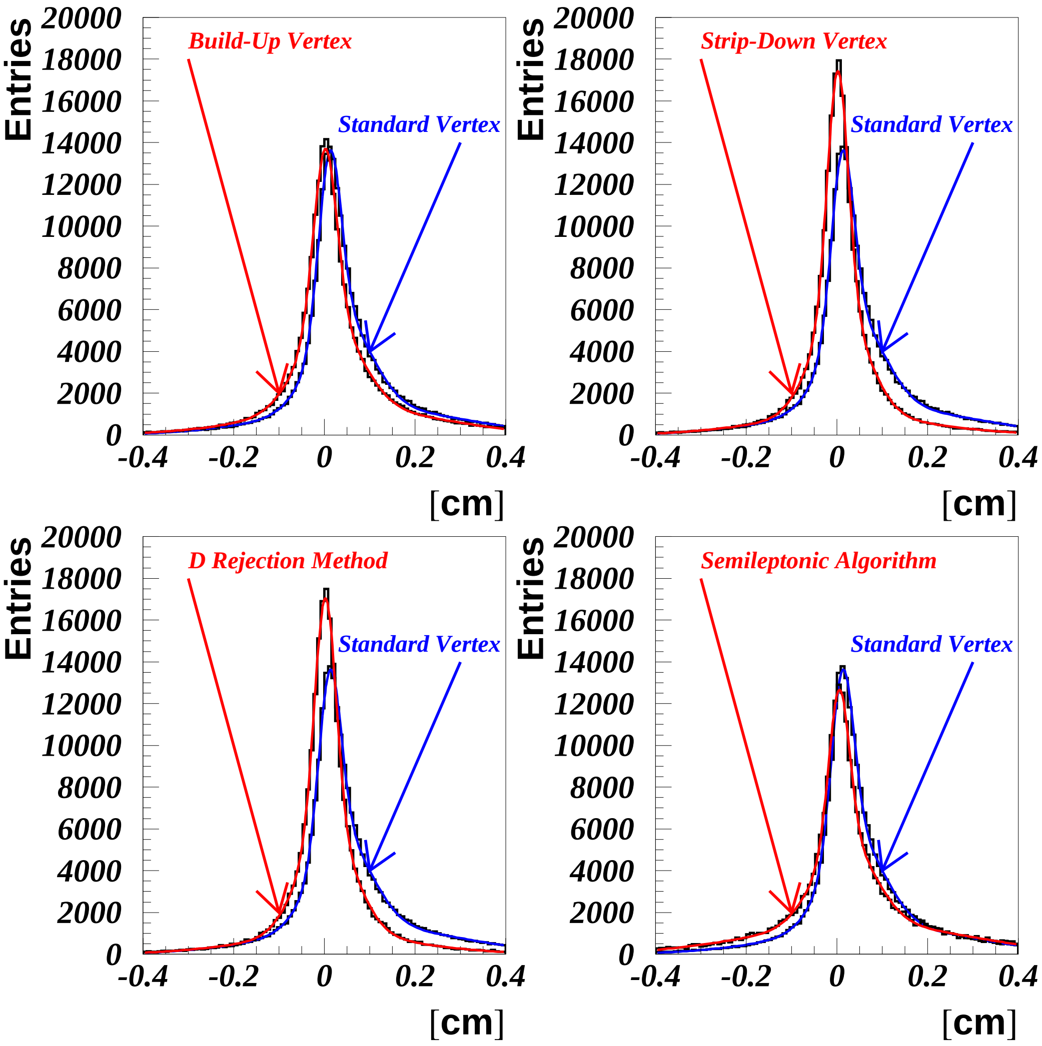

Figure 4.10: *The upper left plot shows a comparison of the Build-Up and the standard algorithm normalized to the standard algorithm distribution. The upper right plots shows the same for the Strip-Down vertex. The bottom plots show the D rejection method and the semileptonic algorithm. It is clearly visible that the Strip-Down vertex nearly removes all forward bias and improves the width of the distribution. For the other algorithms the effects are weaker, but an improvement of the forward bias is also visible.*

This vertex charge distinguishes between charged and neutral b-hadrons and is illustrated in Figure 4.11. Using binomial statistics one can define an estimator for the error of the vertex charge:

$$\sigma_{Q_B} = \sqrt{\sum_{i}^{N_{hem}} P_B(i)(1 - P_B(i))} \tag{4.11}$$

Therefore the vertex charge cannot be estimated with high precision if one or more tracks have a probability P_B around 0.5. The bigger the gap Δ_B between the biggest P_B below 0.5 and the smallest P_B above 0.5 the better the charge reconstruction. Therefore a large hemisphere multiplicity leads to a high error σ_{Q_B}. Another error source which cannot be excluded by a cut on σ_{Q_B} are wrongly reconstructed tracks, so-called 'ghost tracks'. These events are instead coded in a variable called hemisphere quality word.

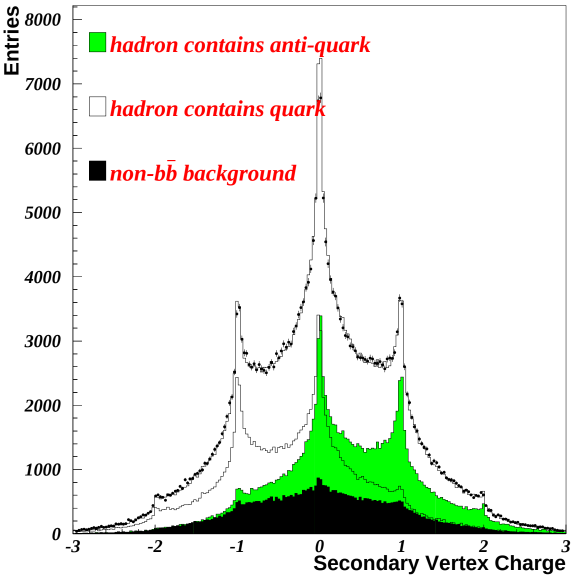

Figure 4.11: *The vertex charge variable comparing data (points) to simulation (histogram) and showing separately the contributions from positively and negatively charged b quarks and from non b events. In addition charged b-hadrons are peaked around ± 1 and neutral b-hadrons are peaked around zero.*

In detail, the inputs of the b species identification network are the following:

- the secondary vertex charge described above to distinguish between charged and neutral b-hadrons
- the binomial error on the vertex charge
- the number of charged pions[9] in the hemisphere. This is most powerful for the case of b-baryons and B_s mesons, which have a higher content of

---

[9] tracks not identified as proton, kaon or lepton are called 'pion'

non-pion particles i.e. neutrons, protons and kaons in comparison to other b species.

- the total energy deposit in the hemisphere by charged and neutral particles scaled by the LEP beam energy. This is sensitive to the presence of b-baryons and B_s mesons, due to the fact that associated neutrons and K^0_L are often not reconstructed in the detector, with the consequence that the total hemisphere energy tends to be smaller compared to B^+ or B^0_d mesons.

- B_s mesons are normally produced with a charged kaon as leading[10] fragmentation particle with a further kaon emerging from the weak decay (the same applies to the associated production of protons with b-baryons). Exploiting this fact, input variables are constructed giving the likelihood for the presence of a leading fragmentation kaon/proton and a kaon/proton weak decay product, for each of the two hypotheses: B_s or b-baryon.

  Utilizing Monte Carlo truth information, normalized track rapidity distributions were parameterized with Gaussians separately for the case of leading fragmentation tracks and b decay products. These were then used to form a weight per reconstructed track which was summed over to give a variable on the hemisphere level.

- the leading fragmentation track can often be neutral (e.g. K^0_S associated with B_s meson production or Λ^0 associated with b-baryon production). An input was therefore constructed based on the presence of a reconstructed K^0_S or Λ^0 in the same way as for charged kaons and protons described above.

- an additional approach, independent of the one mentioned before, is made to get a measure of the probability for the leading fragmentation particle to be a charged kaon. Specifically, the maximum kaon net output from the three tracks with highest rapidity originating from the primary vertex (via the condition TrackNet < 0.5) was used.

- a charge correlation between the leading fragmentation particle charges and the secondary vertex charge provided a flag for the presence of charged b mesons. Specifically, the rapidity-weighted track charge sum over all tracks in the hemisphere is formed and scaled by the measured vertex charge.

In addition, input variables that give no inherent separation power between different b species were included to inform the network of the quality of the other input variables:

---

[10] The term *leading* refers to the neighboring hadron to the b-hadron that emerges from the fragmentation chain. This particle is identified in practice from the fact that it often has the largest rapidity with respect to the b-hadron flight direction.

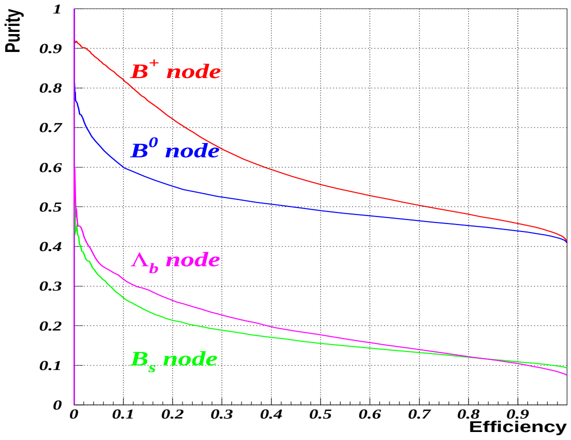

Figure 4.12: *Performance of the SHBN for enriching samples in the various b-hadron types. The plot is based on Monte Carlo making successive cuts in the network output and where 'signal' and 'background' are defined as in Fig. 4.13. The weakest performance is given by the B_s network for which the B_d^0 background is nearly irreducible.*

- the invariant mass of the reconstructed vertex
- the hemisphere quality word already mentioned in Section 4.2.9
- the energy of the b-hadron, to provide information on how hard the fragmentation was and therefore inform the network of how the available energy is expected to be shared between the b-hadron and fragmentation products

The network was trained with an equal number of b-hadrons of each species. The output of the Single Hemisphere B-species identification Networks (SHBN) in both simulation and data is shown in Figure 4.13 and the resulting performance of the SHBN is shown in Figure 4.12. The best performance is achieved for charged b-hadrons, while B_s mesons are hard to separate from the B_d^0 background. The performance can be further improved by the combination of charge correlation information from the opposite hemisphere to form the Both Hemispheres B-species identification Network or BHBN. This network is described in Section 4.2.12.

4.2.11 Flavor Tagging

To reconstruct the flavor content of a hemisphere, i.e. tagging the b quark charge (equivalent to the flavor) of a b jet one can use the charge correlation of a cascade decay of a b quark ($b \to cW^-$ followed by $W^- \to l^-\bar{\nu}$) between the lepton charge and the b quark charge. Unfortunately leptons can be produced with opposite or 'wrong sign' (in D cascade decay). All in all this method is not very efficient because of the small semileptonic branching fraction of about 10%.

In an attempt to use the information available optimally, the BSAURUS approach is tagging the b quark charge by first constructing a track probability and then combining them to give a probability at the hemisphere level.

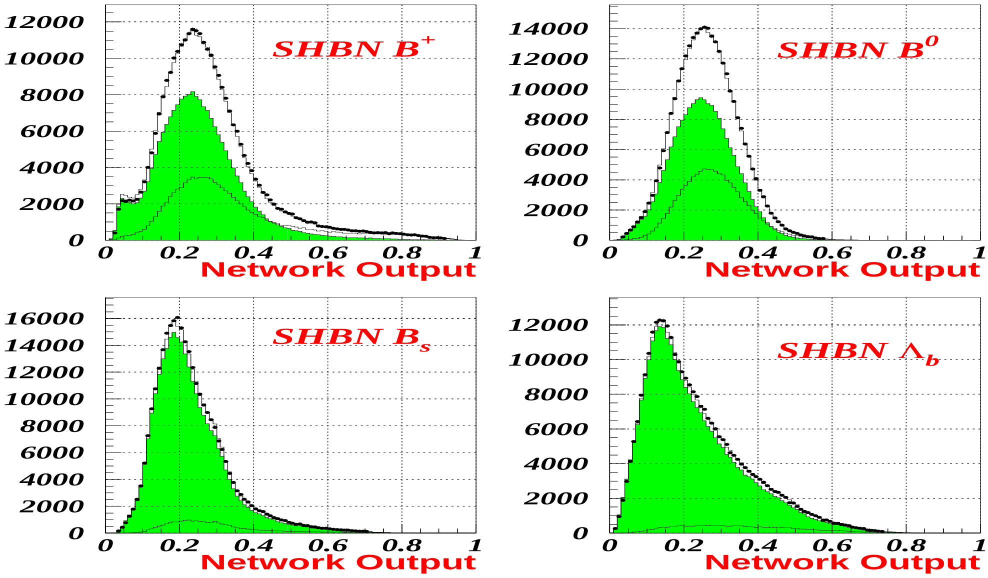

Figure 4.13: *Output of the SHBN network for the* B^+*,* B^0*,* B_s *and b-baryon hypotheses in simulation (histogram) compared to the data (points). The component histograms in each case show the distribution for the hypothesis being considered, i.e. the 'signal' (open histogram), compared to the distribution for everything else, i.e. the 'background' (shaded histogram).*

Several neural networks are trained to tag the underlying quark charge for the cases of a B^+, B^0, B_s or b-baryon in the hemisphere. In addition two sets of such networks are performed, one trained only on tracks originating from the fragmentation process and the other trained only on tracks originating from weak b-hadron decay leading to 8 individual networks. This separation of decay and fragmentation is vital for e.g. oscillation analyses where one has to tag the flavor at decay and production time separately.

These b species dependent flavor networks are combined with the b-hadron identification network (SHBN) described in Section 4.2.10 into a single flavor estimator, i.e. the production flavor.

Track Level Flavor Tagging

The basis of optimal flavor tagging is to construct, at the track level, the conditional probability for the track to have the same charge as the b quark in the b-hadron both at the moment of fragmentation (i.e. production) and at the moment of decay. In addition, these probabilities are constructed separately for each of the b-hadron types B^+, B^0, B_s and b-baryon.

A neural network is used with a target output value of +1(-1) if the track charge is correlated(anti-correlated) with the b quark charge. The discriminating input variables used are the following:

- particle identification variables: kaon, proton and electron net (see [102] for details). Muon classification code (see [104, 105])
- B-D vertex separation variables (only used for decay flavor nets): the track momentum in the b-hadron rest frame, the BDNet output and $\frac{BD-BD(min)}{\Delta BD}$ where BD is the BDNet output, $BD(min)$ is the minimum BDNet value for all tracks in the hemisphere above a TrackNet value of 0.5, and ΔBD is the difference between $BD(max)$ and $BD(min)$
- track level quality variables: the helicity angle in the b rest frame, the track quality word, the TrackNet output and the track energy
- hemisphere-level quality variables: the rapidity gap (see 4.2.9), the hemisphere quality word, the number of tracks passing the standard cuts (Section 4.2.2) and with TrackNet > 0.5, ΔBD as defined above, the secondary vertex mass, the secondary vertex χ^2 probability, b energy divided by beam energy and the error on the vertex charge

In total the track decay flavor network uses 19 input variables, while the track fragmentation uses 14 variables (the B-D information is not valid for fragmentation tracks and the lepton identification is also not used).

Hemisphere Level Decay and Fragmentation Tags

To obtain a flavor tag at the hemisphere level, the conditional track probabilities described above, $P(same\ Q|i)^j$ where $i = B^+,\ B^0,\ B_s$ or b-baryon and $j = fragmentation$ or $decay$, are combined as the likelihood ratio,

$$F(hem)_i^j = \sum_{tracks} \ln\left(\frac{1+P(same\ Q|i)^j}{1-P(same\ Q|i)^j}\right) \cdot Q(track) \tag{4.12}$$

where $Q(track)$ is the track charge. The tracks used in the likelihood sum are tracks with TrackNet > 0.5 for the decay flavor hypothesis and tracks with TrackNet < 0.5 for the fragmentation flavor hypothesis.

The Same Hemisphere Production Flavor Network (SHPN)

The hemisphere flavor tags described in the previous Section form the major input to a further dedicated network, the Same Hemisphere production flavor

Network or SHPN. This network attempts to find the optimal combination of fragmentation and decay flavor, for each b species hypothesis, in order to tag the b-hadron quark flavor at production time. The network is constructed independently of any information from the opposite hemisphere, primarily so that the output can be used for a forward-backward asymmetry ($A_{FB}(b\bar{b})$) measurement [106] incorporating double-hemisphere flavor tag methods.

The 9 input variables used for the network are:

- $F(hem)^{Frag}_{B_s} \cdot P_{SHBN}(B_s)$,[11]
- $(F(hem)^{Dec}_{B^+} - F(hem)^{Frag}_{B^+}) \cdot P_{SHBN}(B^+)$
- $(F(hem)^{Dec}_{bary.} - F(hem)^{Frag}_{bary.}) \cdot P_{SHBN}(bary.)$
- $(F(hem)^{Dec}_{B^0_d} \cdot (1 - 2\sin^2(\frac{\Delta m_d}{2}\tau)) - F(hem)^{Frag}_{B^0_d}) \cdot P_{SHBN}(B^0_d)$, where τ is the reconstructed b lifetime calculated from the decay length of the standard vertex fit and the b-hadron momentum estimate. This construction takes into account the B^0_d oscillation frequency found in the simulation. This is not possible for the case of B_s where the oscillations are so fast that we have essentially a 50%-50% mix of B_s and $\bar{B}_s$.
- The jet charge, which is defined as,

  $$Q_{jet} = \frac{\sum_i Q_i \cdot |\vec{p_i} \cdot \vec{t}|^\kappa}{\sum_i |\vec{p_i} \cdot \vec{t}|^\kappa} \tag{4.13}$$

  where the sum is over all tracks and $\vec{t}$ denotes the thrust axis defined in Section 4.2.3. The optimal choice of the free parameter κ depends on the type of b-hadron under consideration. For this application we choose a range of values $k = 0.3,\ 0.6,\ \infty$, forming individual input variables where the last value corresponds to taking the charge of the track with highest momentum in the hemisphere.
- the vertex charge and its significance

The treatment of quality variables is slightly different for this network compared to other BSAURUS networks. In this case, in order to ensure that the output is inherently symmetric with respect to opposite charges, the quality variables are used to weight the turn-on gradient (or 'temperature') of the sigmoid function used as the network node transfer function. This restriction is especially important for analyses such as measurements of $A_{FB}(b\bar{b})$ which to first order implicitly assume that the charge tag used is symmetric with respect to quark and anti-quark. The quality variables used were:

---

[11] $P_{SHBN}(B_s)$ denotes the B_s mesons probability given by the b-hadron identification network (SHBN)

- the hemisphere quality word
- the hemisphere rapidity gap
- the error on the vertex charge
- the ratio of the b energy to the beam energy

The result of the network is plotted in Figure 4.14 for data and simulation.

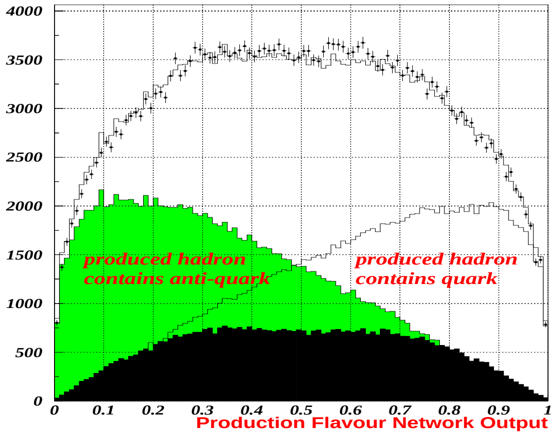

Figure 4.14: *Output of the production flavor net in simulation compared to data. Overlaid are plots from the simulation showing the separation attained for the case where a produced b-hadron contains a b or $\bar{b}$ quark. The black region shows the distribution for the light and charm-quark background.*

4.2.12 Both Hemispheres B-Species Identification Network (BHBN)

The performance of the SHBN (Section 4.2.10) can be further improved by use of the hemisphere flavor tags developed in the previous Section and by the introduction of charge correlation information from the opposite hemisphere to form the Both Hemispheres B-species enrichment Network or BHBN. The variables concerning the flavor of B^0 mesons also contain oscillation information in an explicit form, namely the oscillation probability which is proportional to $\cos(\Delta m_d t)$ where Δm_d is the mixing frequency.

In the same way as for the SHBN, the BHBN is a network with four output target nodes, one for each B type, 16 input nodes and two hidden layers consisting of 20 and 10 nodes respectively. The input variables used, ignoring some details of variable scaling and transformation, are as follows: We define the variable, FL_{opp}, which is a function of the production flavor tag SHPN (see previous Section) in the opposite hemisphere, but scaled by a function in $1-\cos(\theta)$ in an attempt to explicitly account for the b quark forward-backward asymmetry,

$$FL_{opp} = \ln(SHPN(opp)/(1-SHPN(opp))) - \ln((0.5-0.12\cdot\cos(\theta))/(0.5+0.12\cdot\cos(\theta))) \tag{4.14}$$

The discriminating variables used are:

- SHBN probability $P_{SHBN}(i)$ for all four species i
- $F(hem)_i^{Dec} \cdot (FL_{opp} - F(hem)_i^{Frag})$ for all four species i
- $F(hem)_i^{Dec} \cdot (FL_{opp} - F(hem)_i^{Frag}) \cdot P_{SHBN}(i)$ for all four species i
- $FL_{opp} \cdot F(hem)_{B_s}^{Frag}$

Where the $F(hem)$ terms are the hemisphere flavor likelihood ratios described in the previous Section. The quality variables used are:

- the hemisphere quality word
- the hemisphere rapidity gap
- the ratio of the b energy to the LEP beam energy

The outputs of the BHBN are plotted in Figure 4.15. The resulting performance of the network is shown in Figure 4.16.

4.2.13 Reconstructing the b-Hadron Energy by a Neural Network

A neural network was trained to reconstruct the energy of the weakly decaying b-hadron, and a Bayesian interpretation to the network output was applied in order to return a conditional probability density function for the energy on a hemisphere-by-hemisphere basis. In a first step the algorithm took as input the true, i.e. generated, b-hadron energy distribution of a training Monte Carlo sample and transformed it into a uniform distribution in the interval $[0, 1]$. The transformed distribution was then sampled at N equidistant levels, each containing the same number of entries. For each of these levels, a separate output node of a simple feed-forward neural network was trained using back-propagation to the binary classification problem: 'is the true energy value above the level threshold value' vs 'is the true energy value below the level threshold value'. The network outputs were filtered through a sigmoid transfer function and a cubic B-spline fitted through the N net output values. Finally, this spline was interpreted as an estimator of the cumulative probability distribution function (p.d.f.) of the true energy value for a given event and can be regarded as a hemisphere-by-hemisphere unfolding of the input b-hadron energy distribution. The median[12]

---

[12] because it is more robust than the mean

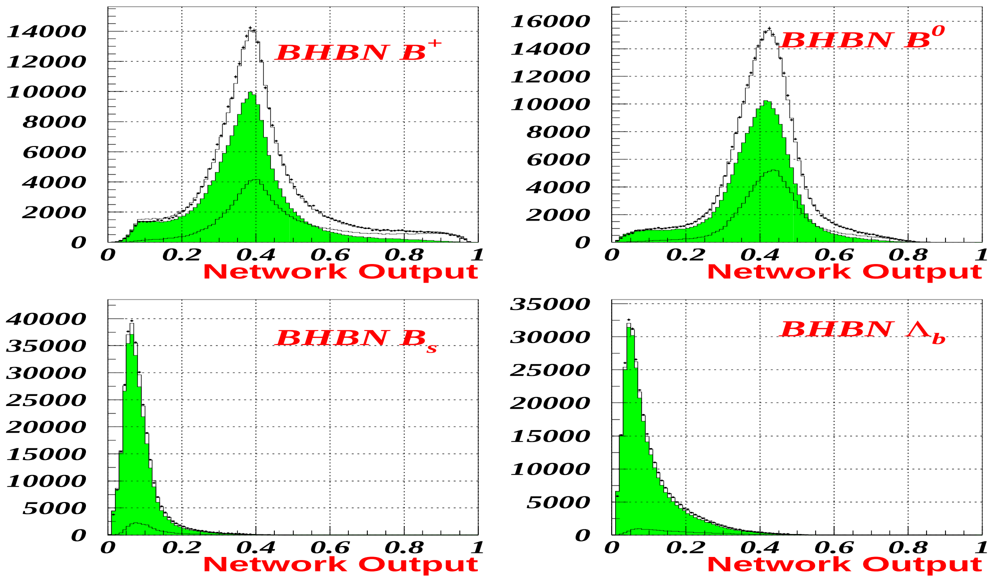

Figure 4.15: *Output of the BHBN network for the* B^+, B^0, B_s *and b-baryon hypotheses in simulation (histogram) compared to the data (points). The component histograms in each case show the distribution for the hypothesis being considered, i.e. the 'signal' (open histogram), compared to the distribution for everything else, i.e. the 'background' (shaded histogram).*

(and associated error) of the estimated p.d.f. then defines the b energy (and error on the energy).

For the network training, sixteen input variables were chosen, which included different estimators of the energy available in the hemisphere together with some measures of the expected quality of such estimators e.g. as given by such quantities as hemisphere track multiplicity and hemisphere reconstructed energy.

The most powerful input variable, with a correlation of 73% to the true b energy , was the corrected energy E_{corr} (see Section 4.2.5). For completeness, the full list of variables input to the energy network were:

- the scaled hemisphere energy x_{hem}
- the raw energy estimate from the rapidity algorithm E_{raw}
- the rapidity mass m_{raw}
- the corrected energy E_{corr}
- the total hemisphere energy normalized by an estimate of the center of mass energy $E_{c.o.m.}$, given by considering the Z^0 to decay into the two-body final

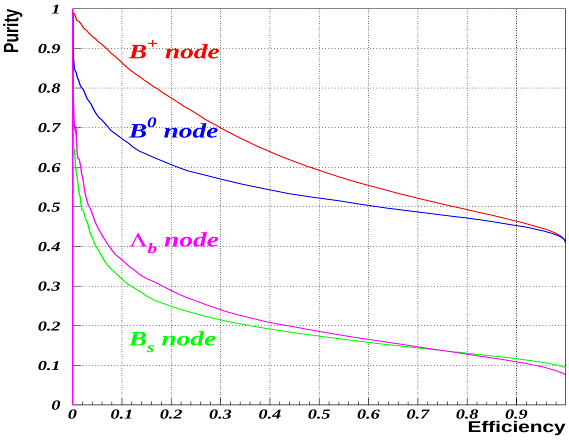

Figure 4.16: *Performance of the BHBN for enriching samples in the various b-hadron types. The plot is based on Monte Carlo making successive cuts in the network output and where 'signal' and 'background' are defined as in Fig. 4.15. The performance is clearly better than for the SHBN.*

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:

$$E_{c.o.m.} = \frac{M^2_{Z^0} - M^2_{recoil} + M^2_{b-jet}}{2 \cdot M_{Z^0}} \tag{4.15}$$

- an estimate of $E_{c.o.m.}$ as given by Eqn. 4.15 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 $E_{c.o.m.}$ as given by Eqn. 4.15 where neutrals from the opposite hemisphere are also included in the formulation of M_{recoil}

- an estimate of the missing momentum 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 the reconstructed primary vertex to the reconstructed secondary vertex and the other given by the vector $\vec{P}_{raw}$.

- the reconstructed secondary vertex mass

- the polar angle of the b candidate momentum vector

- the difference between the number of tracks in the event passing the selection cuts (described in Section 4.2.2) and the number of such tracks that, in addition, pass 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 hemisphere rapidity gap
- the (binomial) error of the vertex charge measurement
- the hemisphere quality word

The training sample was required to pass a combined b-tagging cut of $y_{comb} > 0.5$ and lie in the DELPHI barrel region i.e. $|\cos\theta_{thrust}| < 0.75$. The final b energy resolution from Monte Carlo is plotted in Figure 4.17 and shows a Gaussian peak with non-Gaussian tails.

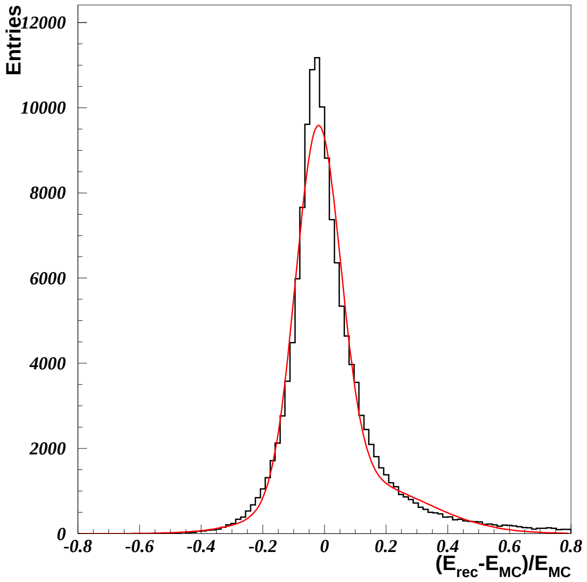

Figure 4.17: *The resolution of the reconstructed b-hadron energy for the Bayesian estimator using neural network techniques. A double Gaussian is fitted to the distribution with widths* $\sigma_1 = 0.073$ *and* $\sigma_2 = 0.21$ *for the inner and outer Gaussian respectively.*

Chapter 5

Studies of a b-Hadron Lifetime Measurement using a Likelihood Method

This chapter contains a study of the measurement of specific b-hadron lifetimes, especially B_s mesons and b-baryons, using a maximum Log-Likelihood method. This method requires a precise modeling and parameterization of the detector acceptance, the resolutions in momentum and decay length and the background. All these ingredients are presented in detail. After that, it is explained how these parameterizations are put together to form the Likelihood formulation. Finally the results of Monte Carlo studies are given which show, that the used method is not robust enough to extract high precision results from data. One possible reason is, that the modeling is not complex enough to account for correlations between measured variables. To have a better model which accounts for these features more Monte Carlo statistics would be necessary.

The first part of the chapter describes briefly the Monte Carlo generator. The sample consists of two parts, a multihadronic event simulation with $Z^0 \rightarrow q\bar{q}$ events and a sample with simulated $Z^0 \rightarrow b\bar{b}$ events only.
The second part of this chapter describes the event selection. The fundamental requirements are a selection of multihadronic events like already mentioned in the previous Chapter, a cut on the b-tagging, which ensures a $b\bar{b}$ sample purity of approximately 94% and a successful run of BSAURUS which ensures a fitted secondary vertex for every event hemisphere. Additional requirements to improve the quality of the sample were the successful application of the Strip-Down algorithm and a cut on the relative expected vertex fit error.
The next Section is dedicated to the modeling of the detector acceptance. The reconstruction efficiency is a function depending strongly the decay time. With smaller decay times the distinction of the secondary and the primary vertex be-

comes more and more difficult. Especially the small lifetime region has therefore to be modeled correctly.
The next Section shortly presents the parameterization of the non-b background which turned out to be not very critical.
Optimal performance can only be reached if the events are classified with respect to their decay length and momentum resolution. Events with good resolution get a higher statistical weight during the fitting routine. The classification is based on the true decay length and the true momentum, both taken from simulation. In addition, the decay length resolution is classified into five momentum classes in order to account for correlations between momentum reconstruction and decay length reconstruction. For the mathematical treatment of the resolution functions in the fitting routine it is necessary to transform the resolution distributions in analytically accessible functions, which is described in this Section as well.
The next Section describes how all the above mentioned ingredients are put together to form a maximum Log-Likelihood fit including the procedure to tag b-hadrons. The results of Monte Carlo studies are presented.
The chapter concludes with a summary and a short discussion of the failure of this approach in reproducing the Monte Carlo lifetimes.

5.1 The Monte Carlo Simulation

For this study, simulated samples were created consisting of 4173514 $Z^0 \to q\bar{q}$ event hemispheres for 1994. The LUND parton shower Monte Carlo program JETSET 7.3 [56] was used where the parameters of the simulator were adapted for Delphi. The detector response has been simulated with DELSIM [55].

Additionally a Monte Carlo sample containing 3066132 $Z^0 \to b\bar{b}$ event hemispheres for 1994 was created to overcome problems with statistical limitations in the parameterization process. The parameters in the simulation which are of special importance for b physics can be found in Table 5.1.

5.2 Event Selection

A lifetime analysis requires a high quality multihadronic sample. To minimize systematic effects due to inexact detector resolution simulation or the modeling of u, d, s and charm production, a relative high b-hadron purity was required. In addition, a class of events was selected which had the best possible decay length resolution.

The first step of the event selection was the separation of multihadronic events ($e^+e^- \to Z^0 \to q\bar{q}$) from various backgrounds like $\tau^+\tau^-$ events, photon-photon

Parameter	**value in Monte Carlo**
B^+ production fraction	41.8%
B^0 production fraction	41.8%
B_s production fraction	7.3%
b-baryon production fraction	9.1%
b-hadron lifetimes	1.6 ps
$BR(B^+ \rightarrow D^-)$	17.7%
$BR(B^0 \rightarrow D^-)$	31.7%
$BR(B_s \rightarrow D^-)$	16.5%
$BR(\Lambda_b \rightarrow D^-)$	12.0%
$BR(B^+ \rightarrow \bar{D}^0)$	69.4%
$BR(B^0 \rightarrow \bar{D}^0)$	55.2%
$BR(B_s \rightarrow \bar{D}^0)$	39.3%
$BR(\Lambda_b \rightarrow \bar{D}^0)$	30.5%
$BR(B^+ \rightarrow \bar{D}_s)$	2.3%
$BR(B^0 \rightarrow \bar{D}_s)$	2.3%
$BR(B_s \rightarrow \bar{D}_s)$	34.0%
$BR(\Lambda_b \rightarrow \bar{D}_s)$	2.2%
$BR(B^+ \rightarrow \Lambda_c)$	5.3%
$BR(B^0 \rightarrow \Lambda_c)$	5.5%
$BR(B_s \rightarrow \Lambda_c)$	5.0%
$BR(\Lambda_b \rightarrow \Lambda_c)$	51.5%
B^+ mass	5.279 GeV/c^2
B^0 mass	5.279 GeV/c^2
B_s mass	5.368 GeV/c^2
Λ_b mass	5.64 GeV/c^2

Table 5.1: DELPHI *Monte Carlo parameter set for b physics.*

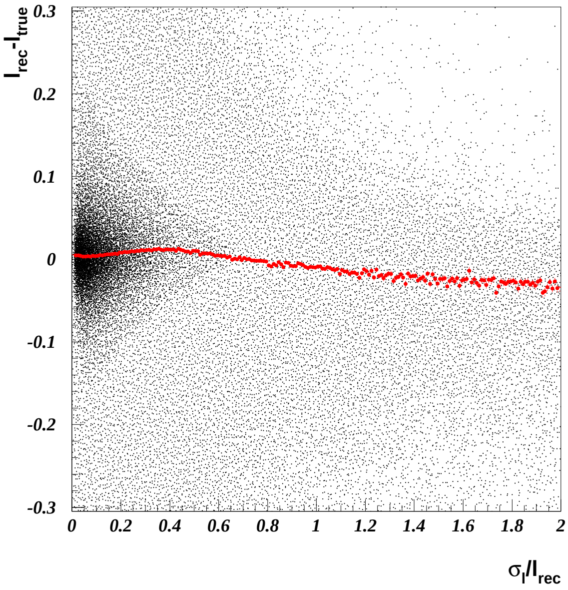

Figure 5.1: *The plot illustrates the resolution of the Strip-Down algorithm versus the relative error. The grey points show the median of the resolution for each bin. It is clearly visible that the resolution shows an increasing bias for relative errors greater than 0.4 which is difficult to be modeled properly.*

events, Bhabha events, beam gas interactions and cosmic rays. The cuts applied to select multihadronic events are similar to the BSAURUS selection described in Section 4.2.1 by requiring at least 5 charged tracks and 12% of the center of mass energy in the event. As outlined in Section 4.2.2 BSAURUS selects in each candidate hemisphere only tracks which pass the standard particle selection criteria. This selection enhances, of course, the probability that BSAURUS does not find a secondary vertex. Hemispheres without a successful secondary vertex fit are removed completely.

In the second step the event jets are reconstructed via the routine LUCLUS [101]. The chosen transverse momentum cutoff value is $d_{join} = 5$ GeV. This essentially restricts the number of jets to about 5. In addition the minimum number of jets has to be two.

The detector has a limited polar acceptance, making all events in very forward/backward direction unusable because a successful vertex reconstruction is mandatory. Therefore events with $(cos(\theta_{thrust}))$ greater than 0.65, where θ_{thrust} is the polar angle of the thrust axis, are rejected to ensure that all selected events have jet axes in the barrel region.

Finally events are selected based on the combined event b-tagging of Section 4.1. For the analysis only event hemispheres are used which have a combined event b-tagging variable larger than 0.3 leading to a b purity of about 94%.

The actual choice of decay length for the fit from different algorithms (see Section 4.2.9) was dictated by optimizing the resolution and minimizing any bias while avoiding underlying substructures from the usage of different resolution classes together. For this purpose the Strip-Down algorithm was chosen because it has the best resolution and the smallest bias. In Figure 5.1 one observes, however, that the resolution of the Strip-Down algorithm shows a bias with increasing relative error. This bias would be difficult to model properly. For this purpose hemispheres are selected only if the relative error was smaller than 0.4. For the analysis only reconstructed decay lengths smaller than 5 cm were accepted to avoid the worst effects of spurious very long decay lengths while still accepting the vast majority of the data available. The number of selected hemispheres after the cuts described in this Section can be found in Table 5.2.

Selection	Number of hemispheres	
	1994 $q\bar{q}$	1994 $b\bar{b}$
Multihadron	4173514	3066132
$\cos(\theta_{thrust})$	2628699	1988411
b-tagging	640888	1581499
Strip-Down vertex	316704	805003
$\sigma_l/l < 0.4$	287232	735313
$l < 5$ cm	287222	735304

Table 5.2: *The number of hemispheres with a successful secondary vertex fit for 1994 simulation. The rows show the number of hemispheres after the multihadron selection (first row), cosine of the thrust axis smaller than 0.65 (second row), the combined event b-tagging cut (third row), the Strip-Down algorithm worked successfully (fourth row), the relative error on the reconstructed decay length smaller than 0.4 (fifth row) and the reconstructed decay length smaller than 5 cm (sixth row).*

5.3 Acceptance Functions

For all b-hadron decays the lifetime distribution is theoretically expected to be an exponential function with mean lifetime τ. However, it is more difficult to separate primary from secondary vertices in b events with very short lifetime, or even recognizing events with very short decay length as b events. Therefore the b-tagging cut (see Section 4.1) and the vertex reconstruction reduces the efficiency for those events. The deviation from the exponential decay is taken from simulation and an acceptance is parametrised accordingly. To take into account differences in the efficiencies for small decay length between different b-hadron species, the acceptance functions have been parametrised separately. The acceptance was parametrised by a spline approximation because an analytic description was not available. However, the spline function did not give a nice approximation for the lifetime region smaller than 2 ps and therefore a linear approximation from bin to bin was used for this region. The acceptance functions $A(t_{true})$ as a function of the true decay time for the four different b-hadron species can be seen in Figure 5.2 together with their parameterizations.

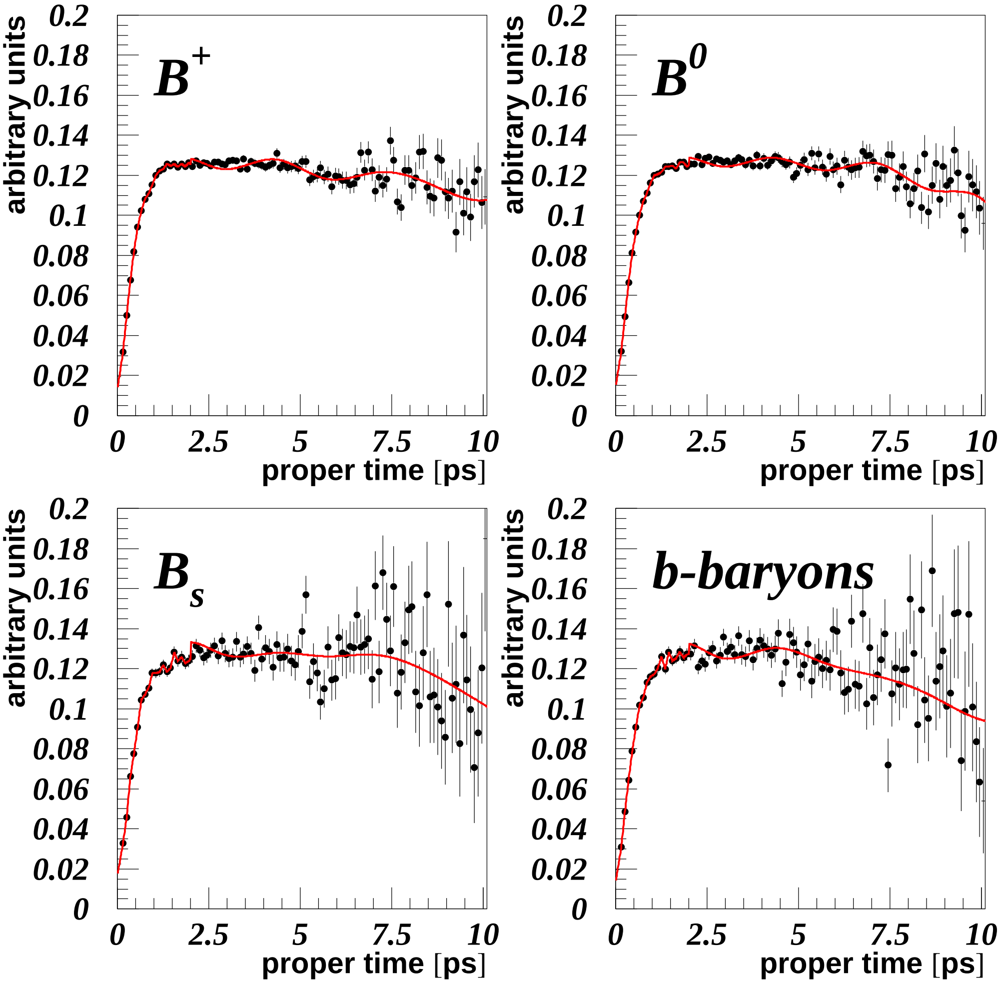

Figure 5.2: *The plots show the acceptance functions for the four different b-hadron species. Superimposed in light grey is the parameterization used for the Likelihood formulation.*

5.4 Background Functions

For charm and light quark events the decay length and momentum distributions had to be modeled. The distributions were taken from the simulation as it is difficult to find a useful analytic description. Also the low statistics make them quite difficult to deal with. The two distributions are plotted in Figure 5.3.

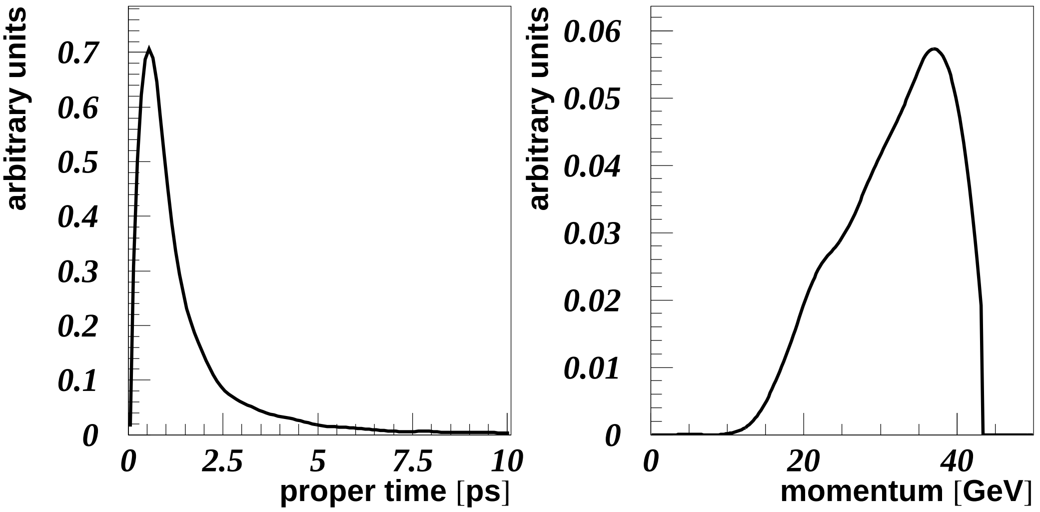

Figure 5.3: *The reconstructed proper time distribution for charm and light quarks (left plot) and their reconstructed momentum distribution (right plot).*

5.5 Modeling the Resolution

5.5.1 Proper Decay Time Reconstruction

The proper decay time was defined by:

$$t = \frac{ml}{cp} \tag{5.1}$$

where m is the mean b-hadron rest mass (in GeV), c the speed of light, p the estimated b-hadron momentum (in GeV) and l the reconstructed decay length. The magnitude of the b candidate momentum vector was fixed by the relationship[1] $E^2 = p^2 + m^2$ where E was the reconstructed b candidate energy. The expected

---

[1] $\hbar = c = 1$

error on the proper time is estimated by error propagation to be:

$$\sigma_t = \sqrt{\left(\frac{m\Delta l}{cp}\right)^2 + \left(\frac{ml\Delta p}{cp^2}\right)^2} \tag{5.2}$$

The error on the decay time has two components which behave differently and that has to be taken into account when modeling the resolution. It motivates the separated treatment of the decay length and momentum resolution in the likelihood fit, including a convolution procedure.
The following Sections are dedicated to the parameterization of the resolution.

5.5.2 Decay Length Resolution

The reconstructed decay length was based on the Strip-Down algorithm as already mentioned with a cut on the relative error σ_l/l at 0.4. The fitting routines for the vertex algorithms had the constraint of reconstructing positive decay lengths only. This complicated the parameterization of the resolution function for the likelihood fit because the resolution becomes a function of the true decay length from simulation. For very small true decay length the resolution becomes more and more asymmetric. This had to be taken into account in the parameterization. For this purpose ten different decay length regions based on the true decay length l_{true} were defined, based on the requirement that the number of event hemispheres in each region was about the same to guarantee that the parameterization did not suffer from small statistics. For this purpose a transformation function was used which was the cumulative probability density function $f(l_{true})$ of the true decay length.

$$f(l_{true}) = s = \int_0^{l_{true}} \mathcal{P}(l)\, dl \tag{5.3}$$

where $\mathcal{P}(l)$ denotes the probability density function of l_{true}. The distribution of the transformed variable s is now flat, and so, to have regions with equal amounts of entries, these regions are chosen to be equidistant bins in s.

To account for possible correlations between the decay length reconstruction and the momentum reconstruction the parameterization was done separately for five regions in the true momentum p_{true}. Again these momentum regions were chosen in a way to have about the same statistics in each region which was made similar to the method mentioned above, but based on the true momentum distribution.

In addition, the resolution has been modeled for the four different b-hadron species considered (B^+, B^0, B_s and b-baryons) separately.

The described classifications divide the sample up into 200 regions that have been modeled by fitting the following asymmetric modified Breit-Wigner function with five parameters P_i to the $l_{rec} - l_{true}$ distributions:

$$f(x, P_i) = P_1 \times \begin{cases} \frac{1}{1+|\frac{x}{P_2}|^{P_3}} & : \quad x \geq 0 \\ \frac{1}{1+|\frac{x}{P_4}|^{P_5}} & : \quad x < 0 \end{cases} \tag{5.4}$$

This rather complicated parameterization had to be chosen to describe properly the non-Gaussian tails in the resolution distributions. The parameter P_1 is used for normalization. Figures 5.4, 5.5, 5.6 and 5.7 show examples of some of the fitted resolution distributions for the ten regions in l_{true} starting at the top left with region 1 and ending on the bottom right with region 10.

Starting from a set of parameters P_i^k, $k \in \{1, \ldots, 10\}$, for every region in decay length a linear interpolation function $P_i(l_{true})$ was calculated to have a smooth 2-dimensional description of the decay length resolution.

The final form of the resolution functions for the different momentum resolution classes p and different b-hadron species s used in the likelihood fit was,

$$\mathcal{R}_s^p(l_{rec} - l_{true}, l_{true}) = \frac{1}{N} \begin{cases} \frac{1}{1+|\frac{l_{rec}-l_{true}}{P_{s2}^p(l_{true})}|^{P_{s3}^p(l_{true})}} & : \quad l_{rec} \geq l_{true} \\ \frac{1}{1+|\frac{l_{rec}-l_{true}}{P_{s4}^p(l_{true})}|^{P_{s5}^p(l_{true})}} & : \quad l_{rec} < l_{true} \end{cases} \tag{5.5}$$

where l_{rec} denotes the reconstructed and l_{true} the true decay length. The factor N is used to preserve the normalization of the resolution function and is calculated numerically. In order to exclude the non-physical region of decay lengths smaller than zero, this region was set to zero and the function renormalized.

5.5.3 Momentum Resolution

The momentum was reconstructed with the algorithm described in Section 4.2.5. Similar to the procedure described in the previous Section the momentum resolution has been modeled. The modeling has been done in ten regions of the true momentum p_{true} for each of the four b-hadrons separately. The regions were chosen to have about the same number of hemispheres in each region in a similar way as for the decay length. Since possible correlations between the momentum resolution and the decay length resolutions have already been accounted for by the division of the decay length resolution into 5 momentum classes, the momentum resolution was not subdivided further. Hence, in total 40 regions have been modeled.

The momentum resolution function was parametrised using a double Gaussian as a function of the difference between reconstructed momentum p_{rec} and true momentum p_{true} using six parameters Q_i in the following way:

$$f(x) = Q_1 \left(Q_2 e^{-\frac{1}{2}\left(\frac{X-Q_3}{Q_4}\right)^2} + (1-Q_2) e^{-\frac{1}{2}\left(\frac{X-Q_5}{Q_6}\right)^2} \right) \tag{5.6}$$

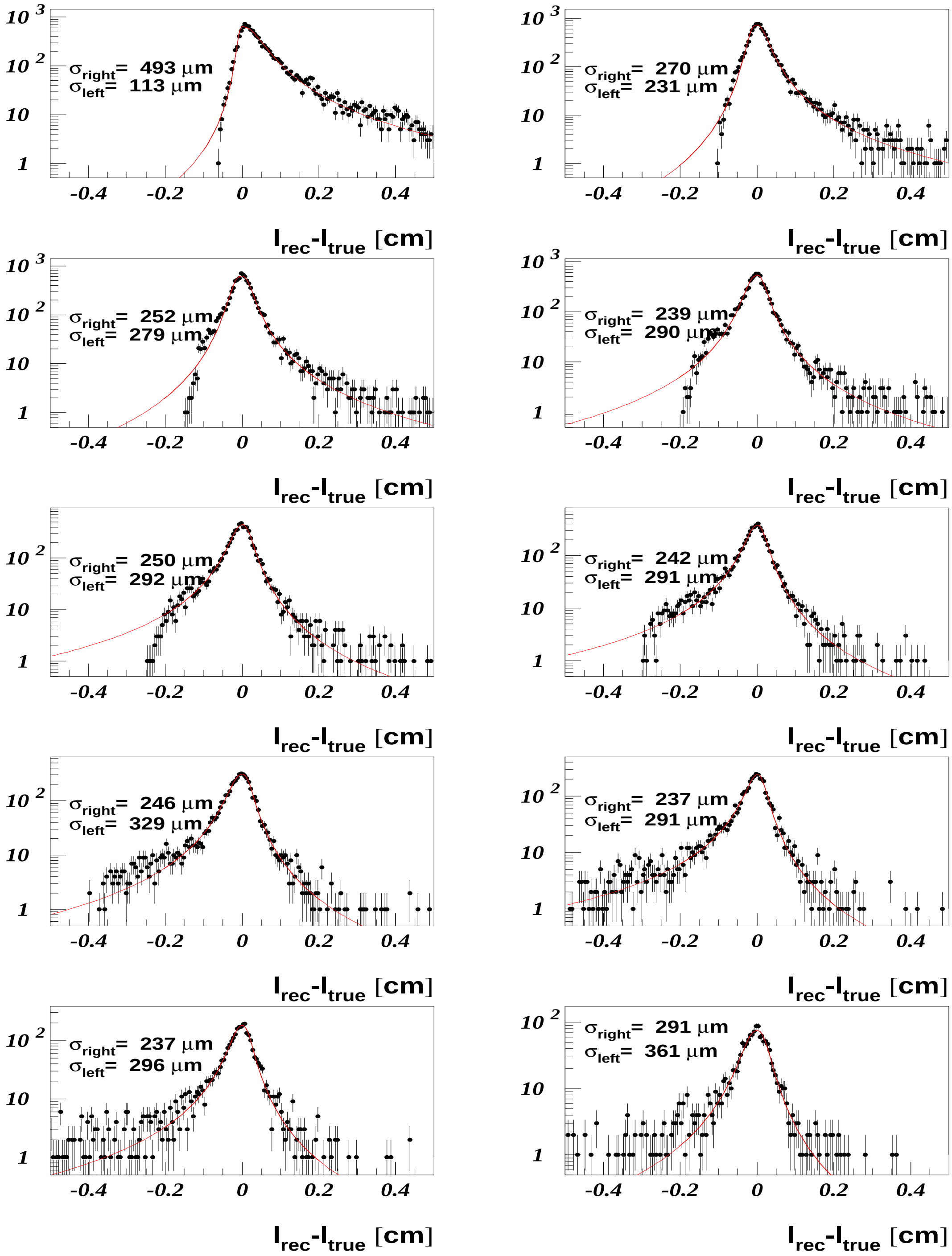

Figure 5.4: *The ten plots illustrate the different shapes of the decay length resolution distribution of B^+-mesons for the ten regions of l_{true} for the momentum class 1. As the vertex algorithm reconstructs only positive decay length the resolution function for the first class in l_{true} is largely biased in forward direction (top left plot). For larger decay length the distribution gets more and more symmetric (bottom right plot). For the plots a logarithmic scale was chosen to show that the function used for the fit gives a good description even in the tails.*

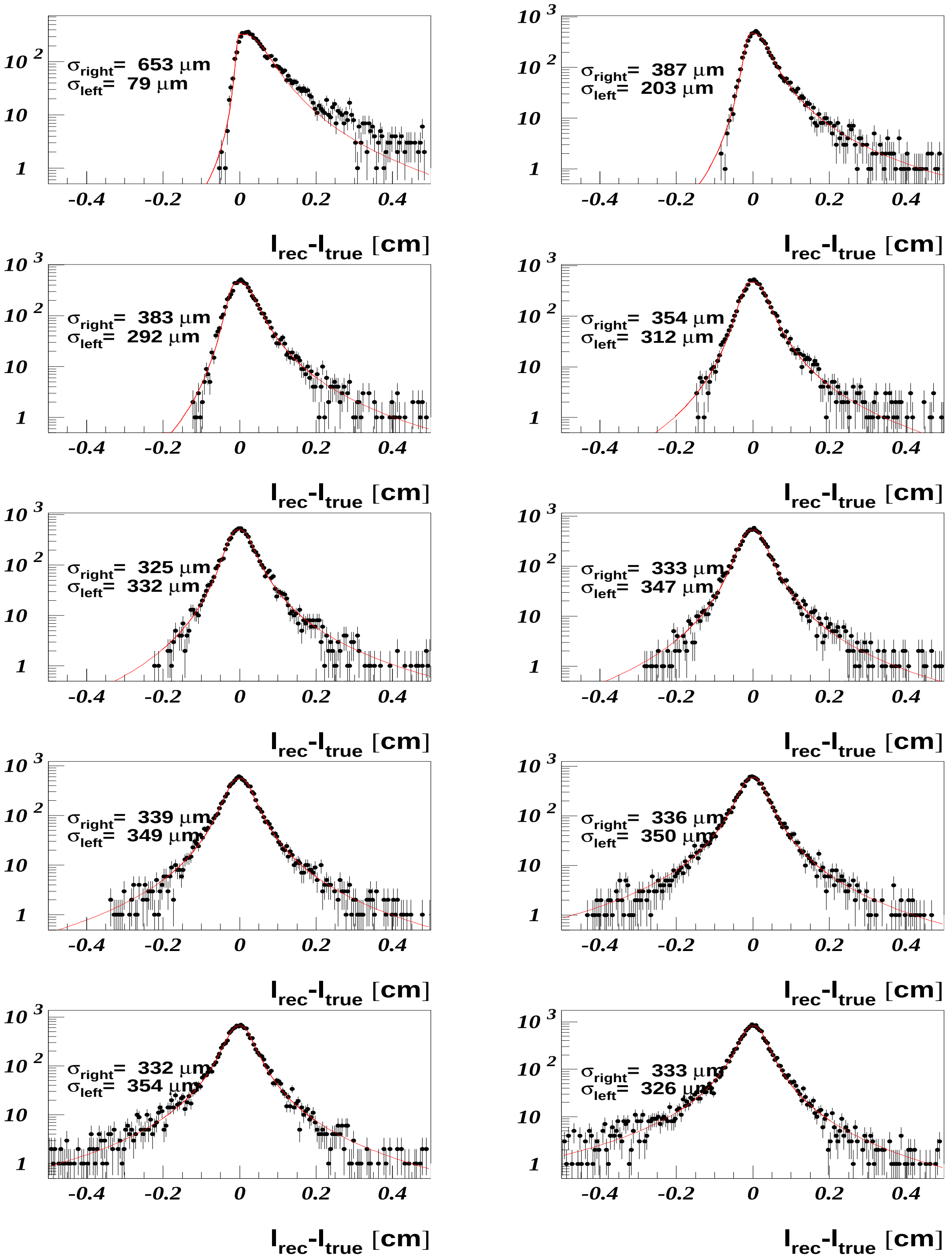

Figure 5.5: *The ten plots illustrate the different shapes of the decay length resolution distribution of B^+-mesons for the ten regions of l_{true} for the momentum class 5.*

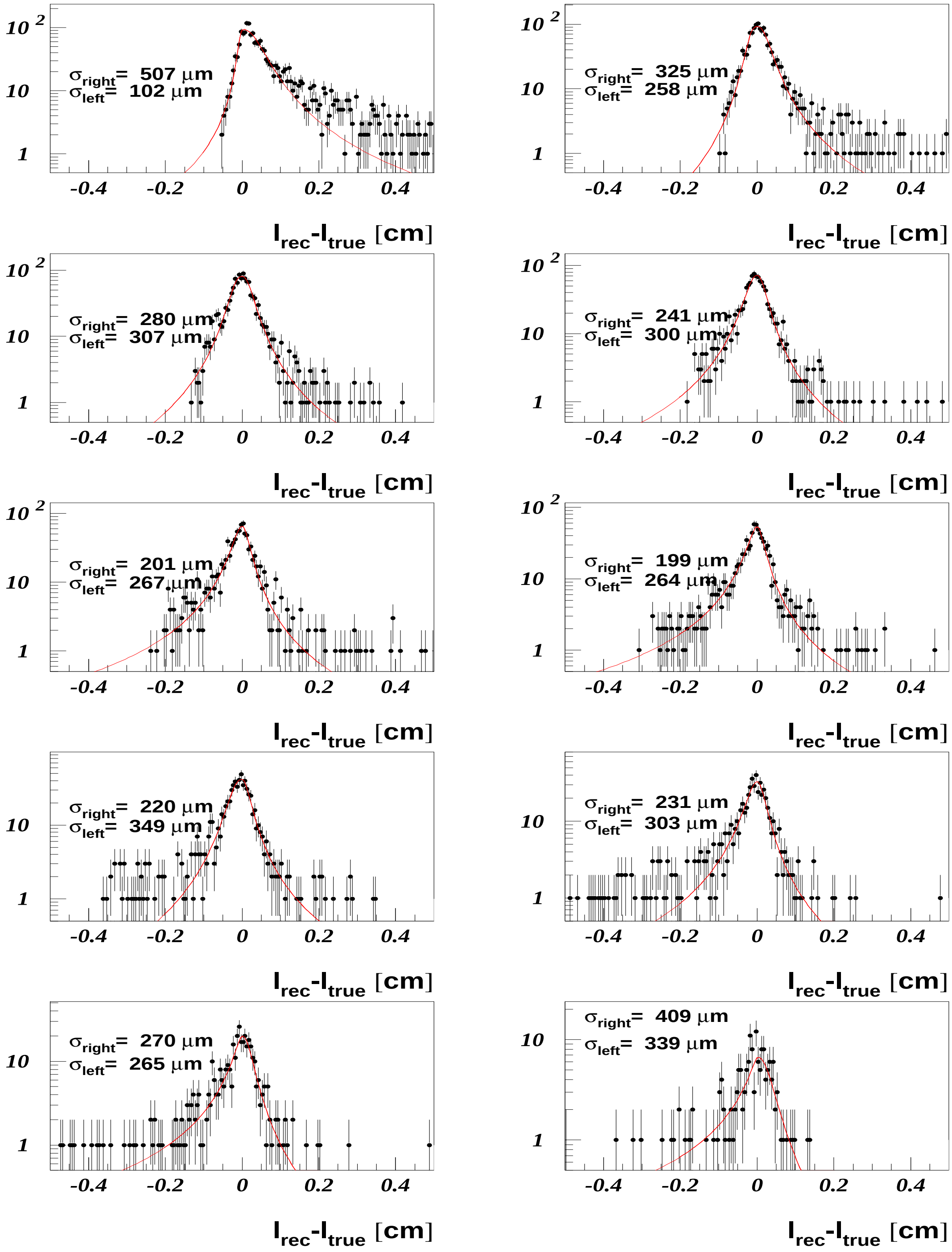

Figure 5.6: *The ten plots illustrate the different shapes of the decay length resolution distribution of b-baryons for the ten regions of* l_{true} *for the momentum class 1.*

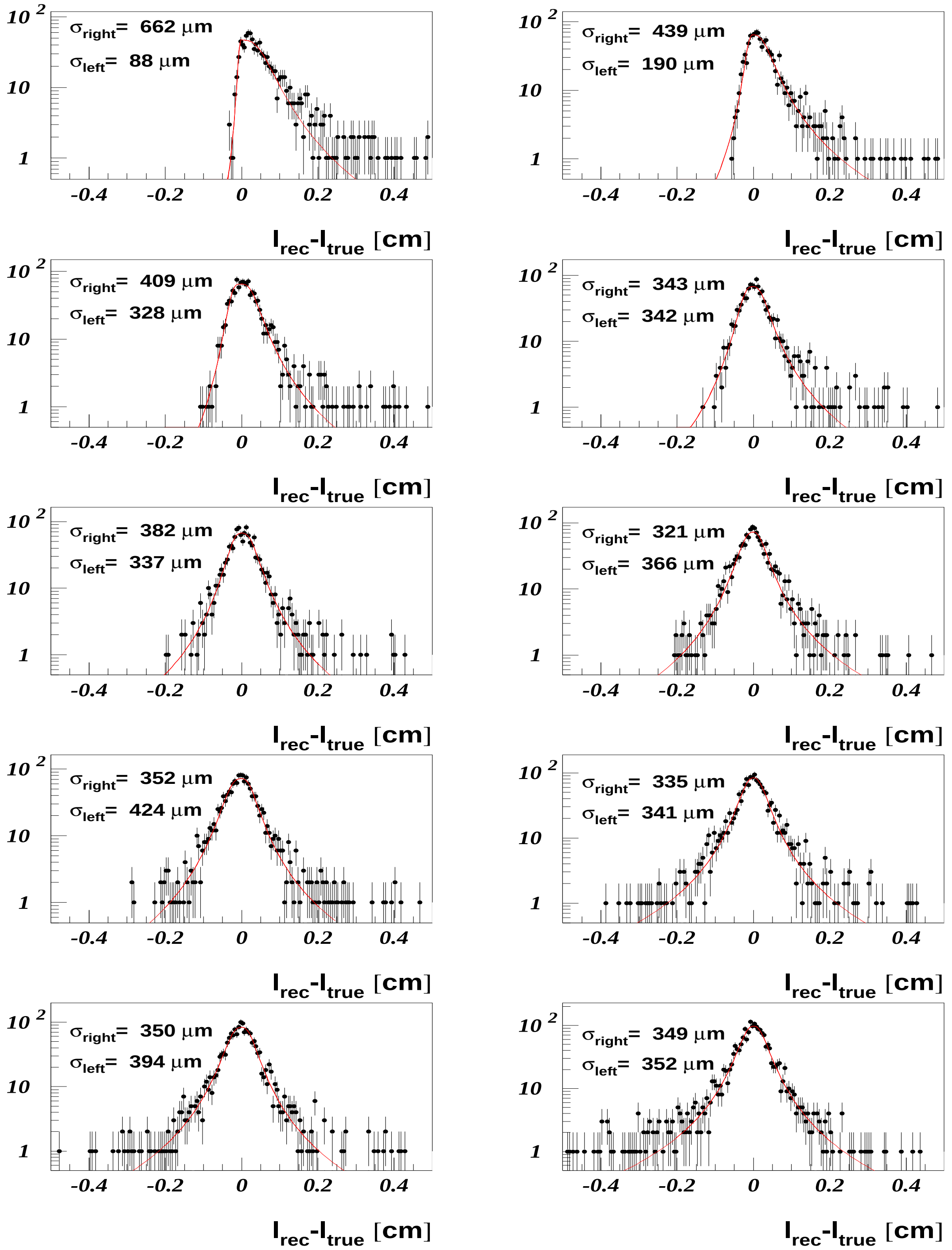

Figure 5.7: *The ten plots illustrate the different shapes of the decay length resolution distribution of b-baryons for the ten regions of* l_{true} *for the momentum class 5.*

The parameters Q_i^k from the ten different regions k are then again used to calculate a linear interpolation function to get a smooth 2-dimensional description of the momentum resolution. Figure 5.8 shows the ten fitted resolution functions for the B^+-meson and Figure 5.9 the same for b-baryons.

5.6 The Likelihood Function

As described in the previous Sections the measurement of the decay length l and momentum p is only possible with finite resolution, parametrised by their resolution functions. The probability function $\mathcal{P}(t_{true})$ of the true decay time for a b-hadron can be written as a function of true decay length and momentum using Equation 5.1,

$$\mathcal{P}(t_{true}) = \frac{1}{\tau_b} e^{-l_{true} m/(\tau_b c p_{true})} \tag{5.7}$$

where τ_b denotes the mean lifetime of the b-hadron. This however is not the combined probability density function $\mathcal{P}(l_{true}, p_{true})$ but it can be easily transformed into it using the conditional probability density function $\mathcal{P}(l_{true}|p_{true})$:

$$\mathcal{P}(l_{true}, p_{true}) = \mathcal{P}(l_{true}|p_{true})\mathcal{P}(p_{true}) = \frac{1}{\tau_b} e^{-l_{true} m/(\tau_b c p_{true})} \mathcal{P}(p_{true}) \frac{m}{c p_{true}} \tag{5.8}$$

where $\mathcal{P}(p_{true})$ is the probability density function for p_{true} taken from simulation and is shown in Figure 5.10. The additional factor m/cp_{true} is a Jacobean factor to guarantee normalization.

The probability $\mathcal{P}(l_{rec}, p_{rec})$ for a b-hadron to be observed with reconstructed decay length l_{rec} and momentum p_{rec} is a convolution of the probability $\mathcal{P}(l_{true}, p_{true})$ multiplied by the acceptance function $A(t_{true})$ by the resolution functions $\mathcal{R}_l$ for the decay length and $\mathcal{R}_p$ for the momentum:

$$\begin{aligned} \mathcal{P}(l_{rec}, p_{rec}) &= \int_0^\infty dl_{true} \int_0^\infty dp_{true} A(t_{true}) \mathcal{P}(l_{true}, p_{true}) \\ &\quad \times \mathcal{R}_l(l_{rec} - l_{true}, l_{true}) \mathcal{R}_p(p_{rec} - p_{true}, p_{true}) \end{aligned} \tag{5.9}$$

5.6.1 The b-Hadron Tagging

In order to be able to distinguish different b-hadron species and to transfer this information into the likelihood formulation a probability density function was constructed based on the b-species identification neural network described in 4.2.12. This p.d.f. was constructed to account for correlations between the network output and the decay time. In order to have a smooth function, cumulative probability density functions are used for both, decay time and network output,

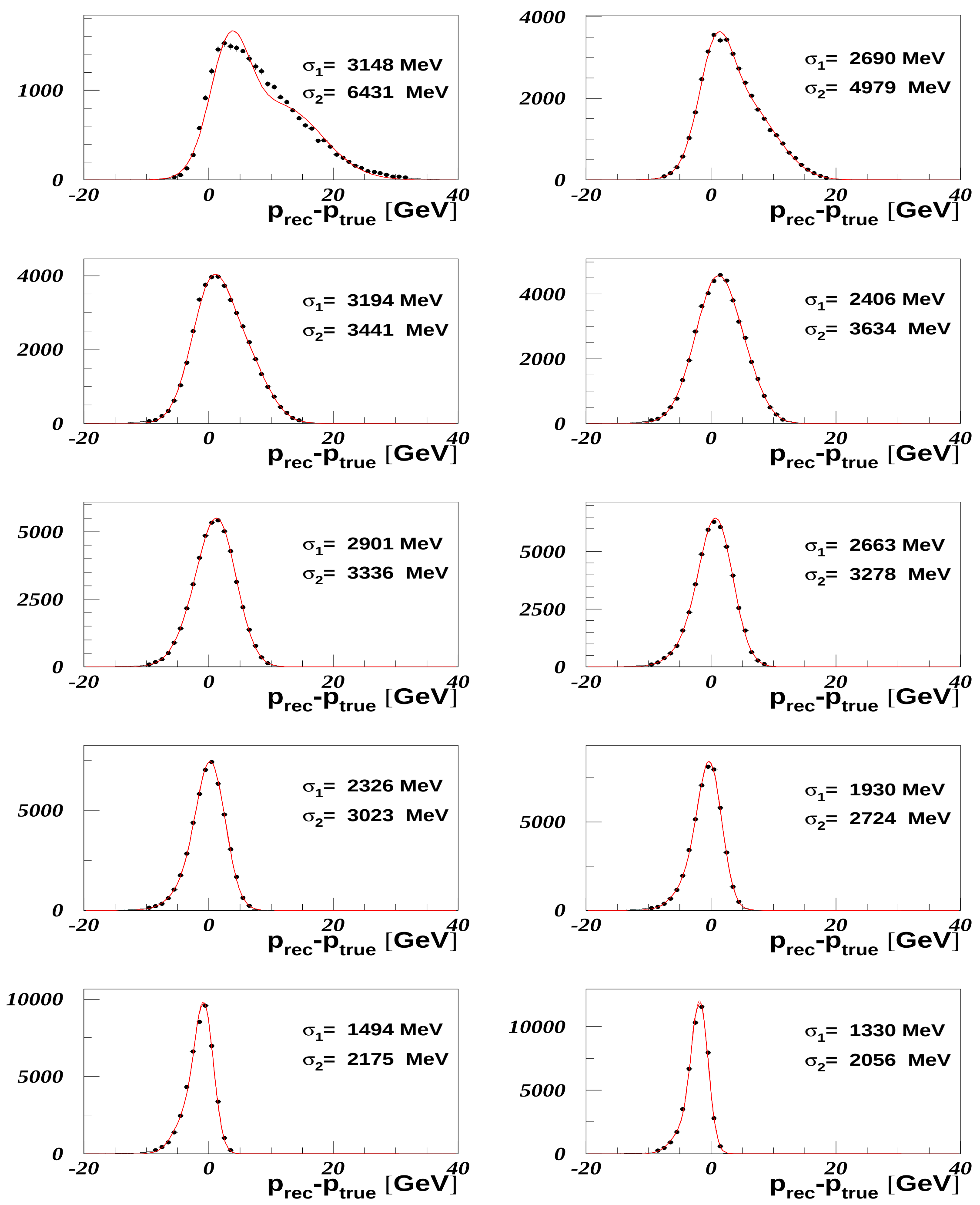

Figure 5.8: *The ten plots illustrate the different shapes of the momentum resolution distribution of B^+-mesons for the ten regions of p_{true}.*

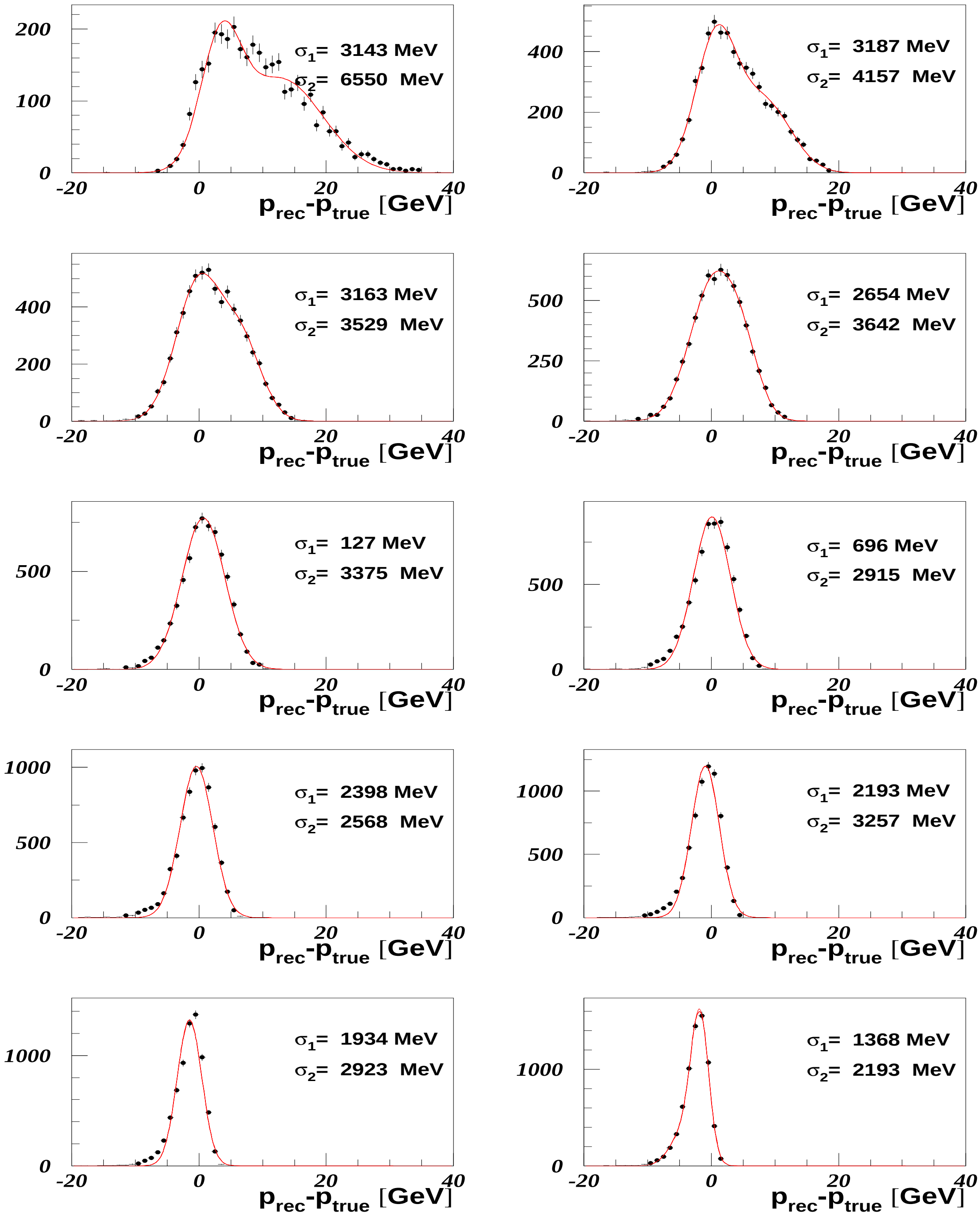

Figure 5.9: *The ten plots illustrate the different shapes of the momentum resolution distribution of b-baryons for the ten regions of* p_{true}.

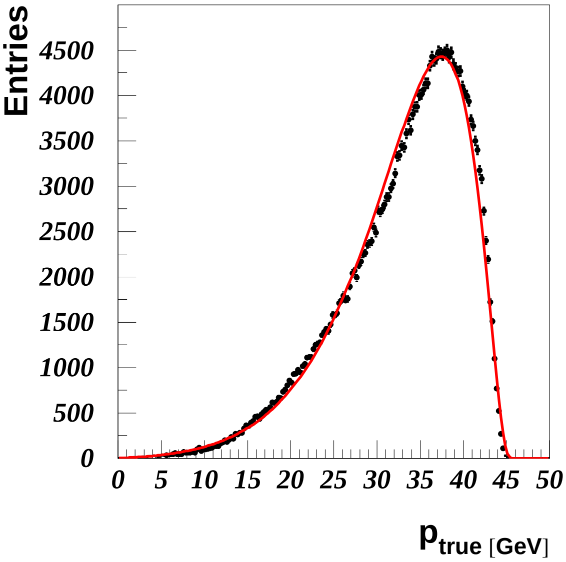

Figure 5.10: *The plot shows the true momentum distribution (points). Overlaid is a fit in order to describe the function analytically (line).*

in the same way as described in Section 5.5.2. Figure 5.11 shows a spline fit of the 2-dimensional p.d.f.'s for the four different output nodes of the neural net. These are normalized along the neural net axis ($f(NN)$). One can clearly see from these plots that the tagging power is a function of lifetime increasing with longer decay times. For the likelihood formulation however only one output node was used. Only one b-hadron lifetime was fitted at a time and so only the corresponding output node was used since this had the biggest power to separate the fitted b-hadron's background, i.e. if the B^+ lifetime was fitted the B^+ output node was used. Figure 5.12 shows spline fits of the p.d.f.'s of the B^+ output node for the four different contributions B^+, B^0, B_s and b-baryons.

Finally all ingredients are put together to form the likelihood function. The probabilities $\mathcal{P}(l_{rec}, p_{rec})$ from Equation 5.9 are modified by the corresponding neural net probability defined above and the convolution integral is calculated for each b-hadron individually. The likelihood function was then the weighted sum of those four probabilities, where the weight was the corresponding b-hadron fraction taken from simulation. Additionally the probability for light and charm quark background are taken directly from the simulation as described in Section 5.4 and are added, weighted with the light and charm quark fraction taken from Monte Carlo. Finally twice of the negative logarithm of the likelihood function is taken to be minimized. The factor two is for convenience reasons only (for further explanation see Section 3.3.2).

In order to guarantee that the normalization of the likelihood function is fixed with respect to the mean lifetime as demanded (see Section 3.3.2), the normalization is calculated numerically in each iteration step of the fitting procedure. All integrals (the integrals for normalization and for the convolution) had to be

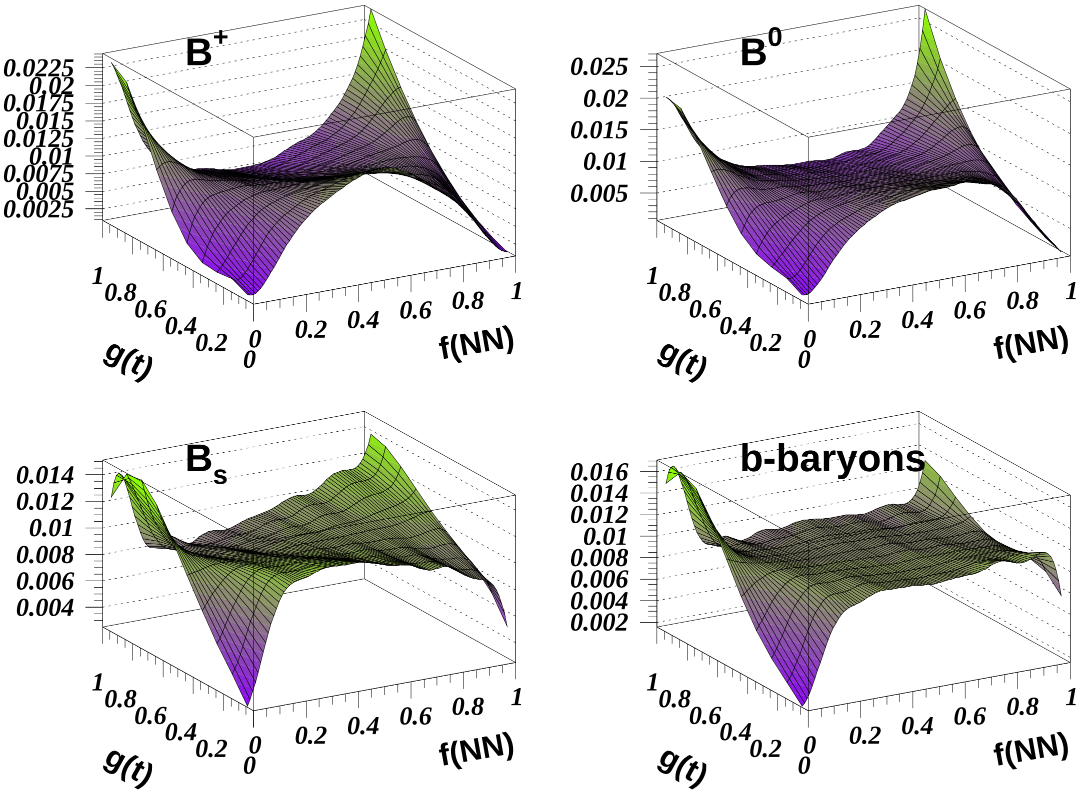

Figure 5.11: *The four plots illustrate the probability function for the four different output nodes of the enhancement network as a function of the cumulative probability density function of the true lifetime vs. the cumulative probability density function of the neural net output variable. One can clearly see that the separation power increases with lifetime.*

performed numerically using techniques described in Section 3.2. Although these techniques are essential in reducing the calculation expense the convolution integrals have to be solved for each event hemisphere and each iteration step. In order to further reduce the number of integrals to be solved, the exponential function $\mathcal{P}(t_{true})$ from Equation 5.7 has been expanded in a Taylor series with respect to the mean lifetime τ_b around the expansion parameter $\tau_0 = 1.55$ ps up to third order:

$$\begin{aligned} e^{-t/\tau_b} \quad \approx \quad & e^{-t/\tau_0} \\ & + e^{-t/\tau_0} \frac{t}{\tau_0^2} (\tau_b - \tau_0) \\ & + e^{-t/\tau_0} \frac{1}{2} \left(\frac{t^2}{\tau_0^4} - \frac{2t}{\tau_0^3} \right) (\tau_b - \tau_0)^2 \end{aligned} \tag{5.10}$$

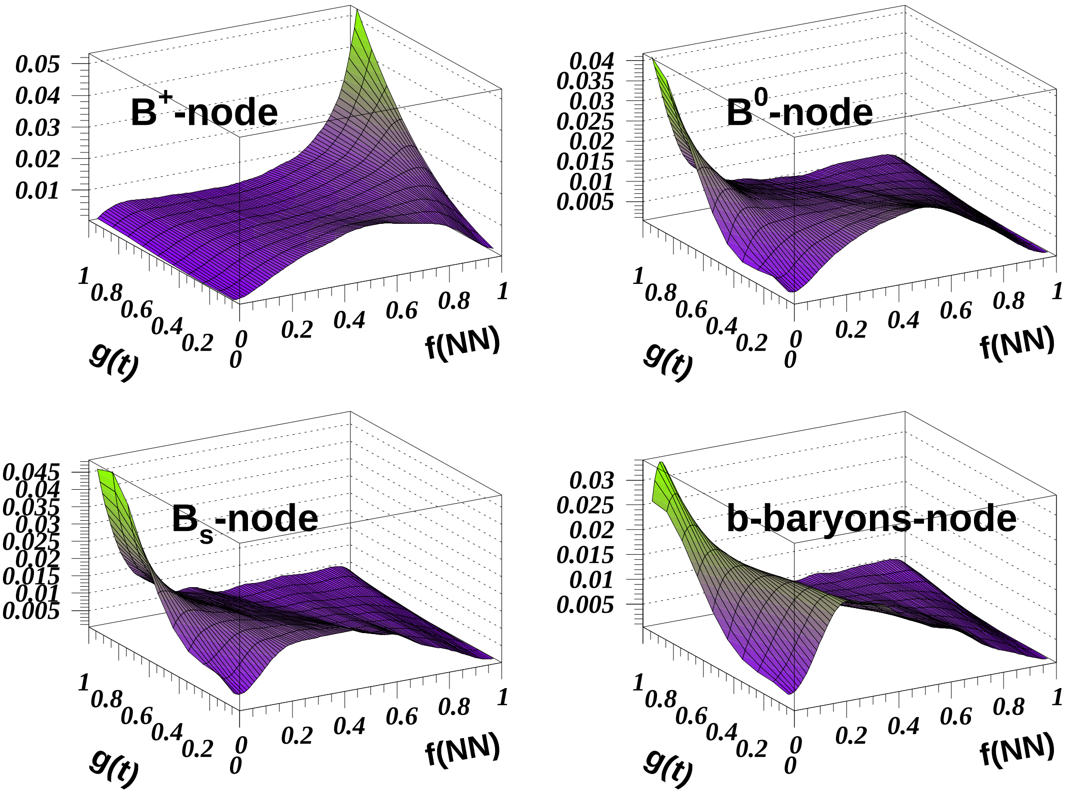

Figure 5.12: *The four plots show the probability function for the B^+ output node of the enhancement network as a function of the cumulative probability density function of the true lifetime vs. the cumulative probability density function of the neural net output variable for the different b-hadron species.*

$$+e^{-t/\tau_0}\frac{1}{6}(\frac{t^3}{\tau_0^6}-\frac{6t^2}{\tau_0^5}+\frac{6t}{\tau_0^4})(\tau_b-\tau_0)^3+\ldots$$

This procedure has the advantage that the convolution integrals have to be calculated only once for each of the four expansion coefficients. The likelihood function then becomes a polynomial of third order in τ_b. During the fit only the normalizations have to be adjusted for each iteration step. In order to check that the order, up to which the expansion is performed, is high enough a test was made to compare the likelihood function with and without Taylor expansion. Figure 5.13 shows the negative log-likelihood function that had to be minimized for the two possibilities for simulation. Around the expected mean lifetime of 1.6 ps the two curves are in good agreement.

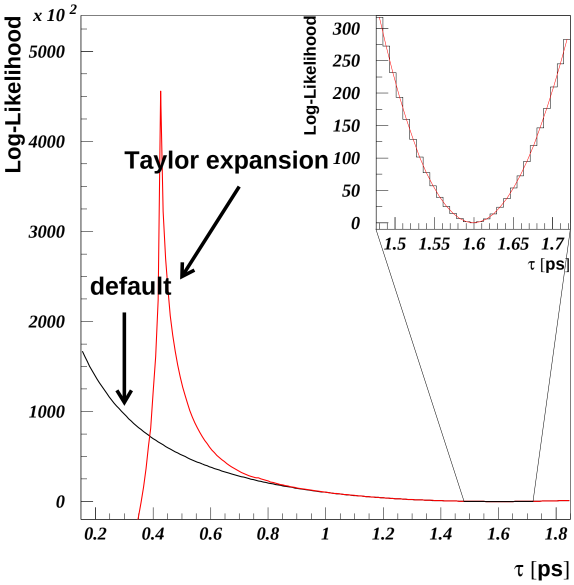

Figure 5.13: *The plot shows the comparison of the log-likelihood function between the formulation using a Taylor expansion of the exponential distribution to the default formulation. In the region of the expected minimum at 1.6 ps the two curves are very close to each other (enlarged section).*

b-hadron	world average [ps]
B^+	1.647
B^0	1.546
B_s	1.464
b-baryons	1.208

Table 5.3: *World average lifetimes from the 'B Lifetime Group'.*

5.6.2 Results from Monte Carlo Tests

In order to test the performance of the method various Monte Carlo tests have been performed to study the robustness of the method. For this purpose simulated events were used as data to check if the method can reproduce the correct mean lifetime which has been generated. Since in simulation all b-hadrons had the same mean lifetime of 1.6 ps this was not really a good challenge to test the method. In order to make the tests more realistic the simulation had to be modified to have different lifetimes for the different b-hadrons. This has been achieved by rejecting events with a certain probability which is based on the true decay time of the b-hadron. Using this method the b-hadron lifetimes had been transformed from 1.6 ps to their current world averages (see Table 5.3) [109]. First the modeling of the acceptance function has been checked. Therefore the true decay length and true momentum was used and no convolution had to be performed. In addition, only 100% pure samples were fitted. The result is shown in Table 5.4.

b-hadron	gen. lifetime [ps]	extracted lifetime [ps]	deviation	gen. lifetime [ps]	extracted lifetime [ps]	deviation
B^+	1.6	1.6059 ± 0.0046	1.3σ	1.647	1.6594 ± 0.0067	1.8σ
B^0	1.6	1.5928 ± 0.0046	1.6σ	1.546	1.5393 ± 0.0063	1.4σ
B_s	1.6	1.6046 ± 0.0122	0.4σ	1.464	1.4738 ± 0.0159	0.6σ
b-baryons	1.6	1.5928 ± 0.0117	0.6σ	1.208	1.2023 ± 0.0156	0.4σ

Table 5.4: *Results of the test using true information for generated lifetimes of 1.6 ps and results with generated world average lifetimes.*

b-hadron	gen. lifetime [ps]	extracted lifetime [ps]	deviation	gen. lifetime [ps]	extracted lifetime [ps]	deviation
B^+	1.6	1.5975 ± 0.0050	0.5σ	1.647	1.6394 ± 0.0071	1.1σ
B^0	1.6	1.5874 ± 0.0050	2.5σ	1.546	1.5258 ± 0.0068	3.0σ
B_s	1.6	1.6145 ± 0.0135	1.1σ	1.464	1.4722 ± 0.0176	0.5σ
b-baryons	1.6	1.5153 ± 0.0370	2.3σ	1.208	1.1737 ± 0.0364	0.9σ

Table 5.5: *Results of the test using the resolution functions to smear the true information and performing the convolution integrals.*

In a second step the resolution functions described in Section 5.5 were used to smear the true decay length and true momentum. This test should demonstrate, if the convolution was performed in a correct way to get back the generated lifetimes. To avoid complications with contributions from other b-hadrons again only 100% pure samples were fitted. The result is shown in Table 5.5.

Next the fit was performed on reconstructed Monte Carlo events to see if the resolution modeling was correct, or if some hidden correlations have not been taken into account. The result of this test is shown in Table 5.6. The above three tests have been repeated for the case of using natural admixture of b-hadrons in the samples while taking the fraction from the simulation. The results of those tests can be found in Tables 5.7, 5.8 and 5.9. As one can see from these tests, the numerical treatment of the acceptance functions, the resolutions and the convolution works quite well. However, when using reconstructed decay length and momentum the method is not able to recover the correct lifetimes which were put in. Additional tests showed a strong dependence of the result with respect to the acceptance modeling. Using a spline fit or an adequate parameterization for the low lifetime region resulted in a deviation from the expected lifetimes of several standard deviations. Since it is known that modeling the region at low proper time in the Monte Carlo simulation is a particularly complex task and far

b-hadron	gen. lifetime [ps]	extracted lifetime [ps]	deviation	gen. lifetime [ps]	extracted lifetime [ps]	deviation
B^+	1.6	1.5865 ± 0.0049	2.8σ	1.647	1.6389 ± 0.0072	1.1σ
B^0	1.6	1.5776 ± 0.0049	4.5σ	1.546	1.5244 ± 0.0068	3.2σ
B_s	1.6	1.5991 ± 0.0133	0.1σ	1.464	1.4780 ± 0.0175	0.8σ
b-baryons	1.6	1.5816 ± 0.0126	1.5σ	1.208	1.2073 ± 0.0140	0.1σ

Table 5.6: *Results of the test using reconstructed decay length and momentum from the Monte Carlo simulation.*

b-hadron	gen. lifetime [ps]	extracted lifetime [ps]	deviation	gen. lifetime [ps]	extracted lifetime [ps]	deviation
B^+	1.6	1.5994 ± 0.0069	0.1σ	1.647	1.6540 ± 0.0096	0.7σ
B^0	1.6	1.5972 ± 0.0070	0.4σ	1.546	1.5516 ± 0.0097	0.6σ
B_s	1.6	1.6017 ± 0.0512	0.0σ	1.464	1.5169 ± 0.0714	0.7σ
b-baryons	1.6	1.5798 ± 0.0476	0.4σ	1.208	1.2394 ± 0.0628	0.5σ

Table 5.7: *Results of the test using true information and a natural admixture of b-hadrons.*

from being perfect one cannot expect to get meaningful results from a data fit having a method with that low level of robustness.

5.7 Summary and Discussion

A study was performed for the inclusive measurement of b-hadron lifetimes using a Likelihood method.

After selecting b hadron candidates with a purity of about 94%, candidates with good resolution were selected. All inputs for the likelihood formulation of the problem were parameterized. Then all ingredients were put together to check the method on a Monte Carlo sample, which in one case had the lifetimes of the generator at 1.6 ps and in the other case four different lifetimes. For this purpose three different checks were performed on 100% pure samples which represented three stages of difficulty to extract the lifetimes. The first check was performed on the true information for decay length and momentum. Therefore no convolution integrals had to be solved. This check reproduced the lifetimes quite well. The second check was also performed on the true information but this time it has

b-hadron	gen. lifetime [ps]	extracted lifetime [ps]	deviation	gen. lifetime [ps]	extracted lifetime [ps]	deviation
B^+	1.6	1.5851 ± 0.0075	2.0σ	1.647	1.6357 ± 0.0104	1.1σ
B^0	1.6	1.5844 ± 0.0076	2.1σ	1.546	1.5333 ± 0.0106	1.2σ
B_s	1.6	1.5077 ± 0.0582	1.6σ	1.464	1.3709 ± 0.0721	1.3σ
b-baryons	1.6	1.4641 ± 0.0538	2.5σ	1.208	1.2098 ± 0.0416	0.0σ

Table 5.8: *Results of the test using the resolution functions to smear the true information and performing the convolution integrals with a natural admixture of b-hadrons.*

b-hadron	gen. lifetime [ps]	extracted lifetime [ps]	deviation	gen. lifetime [ps]	extracted lifetime [ps]	deviation
B^+	1.6	1.5635 ± 0.0073	5.0σ	1.647	1.6158 ± 0.0102	3.1σ
B^0	1.6	1.5601 ± 0.0074	5.4σ	1.546	1.5155 ± 0.0103	3.0σ
B_s	1.6	1.3890 ± 0.0405	5.2σ	1.464	1.3372 ± 0.0583	2.2σ
b-baryons	1.6	1.3856 ± 0.0423	5.1σ	1.208	1.1985 ± 0.0380	0.0σ

Table 5.9: *Results of the test using using reconstructed decay length and momentum from the Monte Carlo simulation with a natural admixture of b-hadrons.*

been smeared according to the parameterized resolution functions. Also by this test the lifetimes had been reproduced quite well. Finally the third check was performed using reconstructed decay length and momentum. This time the result from the fit differed from the expected result by several standard deviations. The three checks were repeated using natural b-hadron composition showing basically the same behavior in the results.

This leads to the conclusion that the model and parameterization is not good enough to perform a high precision measurement on b-hadron lifetimes. There might be some correlations to variables which have not been considered yet. For example it is possible, that there are different underlying resolution classes which have not been taken into account explicitely. This would explain that the check using the smeared true information did not show a problem. In this test the smeared decay length and momentum had been modeled exactly without any hidden correlation. Several other tests showed that the likelihood method is very sensitive to the exact modeling of the acceptance functions. Since it is a known fact that the Monte Carlo simulation for the low lifetime region is a rather bad model one cannot expect to extract reliable results on data using a method which is that sensitive to small changes in the parameterization.

Chapter 6

Measurement of the Lifetimes of B^+ and B^0 Mesons

This chapter provides an analysis for the measurement of the lifetimes of B^+ and B^0 mesons. All ingredients used in this analysis are described in detail, i.e. event selection, sample classification, decay time definition and Monte Carlo reweighting. After that the fitting method (χ^2 *minimization*) and the choice of the working point is explained. Finally the results are presented and the details of the considered systematic error sources are given. The chapter finishes with a short summary.

The first part of the chapter describes briefly the Monte Carlo generator. The Monte Carlo sample consists of two parts, a multihadronic event simulation with $Z^0 \to q\bar{q}$ events and a sample with simulated $Z^0 \to b\bar{b}$ events only.

The event selection of the data sample is described in the second part of the chapter. The fundamental requirements are a selection of multihadronic events like already mentioned in Chapter 4, a cut on the b-tagging, which ensures a $b\bar{b}$ sample purity of approximately 94% and a successful run of BSAURUS which ensures a successful fitted secondary vertex for every event hemisphere. In total 421913 event hemispheres have been selected in the data sample from 1994 to 1995 by these pre-cuts.

From the four different advanced secondary vertex algorithms presented in the Chapter 4 a special selection had to be made for each hemisphere. This selected decay length and the energy estimator from the Bayesian approach were used to calculate the proper decay time per hemisphere.

In order to improve the agreement with the data of the Monte Carlo and to account for known discrepancies in the Monte Carlo generator with recent experimental measurements, weights were constructed. These weights which have been applied to all Monte Carlo hemispheres before the fitting procedure are

b-hadron	MC prod. fraction [%]	MC lifetime [ps]
B^+	41.8	1.6
B^0	41.8	1.6
B_s	7.3	1.6
b-baryons	9.1	1.6

Table 6.1: *The Monte Carlo production fractions and lifetimes for b-flavored particles used in the Monte Carlo generator*

presented afterwards.

The division into two enhanced samples is described in Section five leading to 69310 selected hemispheres in total for the years 1994 to 1995.

The results are extracted in terms of a binned χ^2 minimization fit. The detailed formulation of the fit is presented in the sixth Section.

Section seven describes in detail the choice of the working point including the presentation of stability plots around it. The section also provides all systematic checks and studies which were necessary to guarantee a robust fit result. Further the section contains a description of the method to combine the two separate measurements for 1994 and 1995 data to give one single result.

In section eight all systematic uncertainties considered within this analysis are described in detail. It finally contains the full systematic error breakdown for 1994 and 1995 separately.

The last section gives a short summary of the chapter.

6.1 The Monte Carlo Simulation

For the analysis simulated samples were created consisting of 4173514 $Z^0 \to q\bar{q}$ event hemispheres for 1994 and 1565386 event hemispheres for 1995. For this the LUND parton shower Monte Carlo program JETSET 7.3 [56] was used were the parameters of the simulator were adapted for DELPHI. The detector response has been simulated with DELSIM [55].

Additionally a Monte Carlo sample containing 3066132 $Z^0 \to b\bar{b}$ event hemispheres for 1994 and 824924 event hemispheres for 1995 was created to restrict the statistical limitation of the analysis to real data only. The simulated events are fed into the same analysis chain than the real data events. The relative b-hadron production fractions and lifetimes used for the generator have their origins in theoretical predictions and can be found in table 6.1.

Selection	Number of hemispheres					
	1994 data	1995 data	1994 $q\bar{q}$	1994 $b\bar{b}$	1995 $q\bar{q}$	1995 $b\bar{b}$
Multihadron	1922532	927679	4173514	3066132	1565386	824924
$\cos(\theta_{thrust})$	1199774	577264	2628699	1988411	982705	535300
b-tagging	285088	136825	640888	1581499	239061	422304

Table 6.2: *The number of hemispheres with a successful secondary vertex fit for data and simulation. The rows show the number of hemispheres after the multihadron selection (first row), cosine of the thrust axis smaller than 0.65 (second row) and the combined event b-tagging cut (third row).*

6.2 Event Selection

A lifetime analysis is only possible on a well selected, high quality multihadronic sample. The selection conditions imposed on the data samples used for the lifetime fits were motivated by the wish to minimize the total error on the final results. Systematic error contributions due to the inexact simulation of detector resolutions and the physics modeling of u, d, s and charm production mean that relatively high b-hadron purities were required while still keeping the selection efficiency at an acceptable level as far as the statistical error is concerned.

The first step of the event selection is the separation of multihadronic events ($e^+e^- \rightarrow Z^0 \rightarrow q\bar{q}$) from various backgrounds. The background sources are widely spread, e.g. $Z^0 \rightarrow \tau^+\tau^-$ events, photon-photon events, Bhabha events beam gas interaction events and cosmic rays. Multihadronic events are selected similar to the BSAURUS selection described in Section 4.2.1 by requiring at least 5 charged tracks and 12% of the center of mass energy in the event. As outlined in Section 4.2.2 BSAURUS selects in each candidate hemisphere only tracks which pass the standard particle selection criteria. This selection enhances of course the probability that BSAURUS does not find a secondary vertex. Hemispheres without a successful secondary vertex fit are removed completely.

In the second step the event jets are reconstructed via the routine LUCLUS [101]. The chosen transverse momentum cutoff value is $d_{join} = 5$ GeV. This essentially restricts the number of jets to about 5. In addition the minimum number of jets has to be two.

As described in chapter two the detector has a limited polar acceptance, making all events in very forward/backward direction unusable because a successful vertex reconstruction is mandatory. Therefore events with ($cos(\theta_{thrust})$) greater than 0.65 are rejected to ensure that all events selected have jet axes which fall into the efficient barrel region.

The final event selection cut is based on the combined event b-tagging of

Section 4.1. For the analysis only event hemispheres are used which have a combined event b-tagging variable greater than 0.3.

After all the selection criteria above the sample b purity is about 94%. In table 6.2 the numbers of hemispheres with a successful secondary vertex fit for each subsequent selection is listed for the years 1994 and 1995 and for the Monte Carlo simulation samples.

6.3 Proper Decay Time Reconstruction

The proper decay time t was defined by the following relationship,

$$t = \frac{ml}{cp} \tag{6.1}$$

where m is the mean b-hadron rest mass which was taken to be 5.2789 Gev/c^2 [107], c the speed of light, p the estimated b-hadron momentum and l the reconstructed decay length. The magnitude of the b candidate momentum vector was fixed by the relationship $E^2 = p^2 + m^2$ where E was the reconstructed B candidate energy. Only hemispheres were accepted, were the reconstructed energy was greater than 10 GeV. The estimator of the B candidate energy was the Bayesian neural network estimator described in Section 4.2.13.

The actual choice of decay length for the decay time calculation from the different algorithms (see Section 4.2.9) was dictated by optimizing the resolution and minimizing any bias while still retaining the best possible efficiency. The choice from one of the four algorithms potentially available, was made in the following, subsequent way:

- the Strip-Down method was chosen if the algorithm worked successfully and had a decay length error smaller than 1 mm,
- next, if the hemisphere failed the criterias above, the D-rejection method was used, if successful, and had a decay length error smaller than 1 mm,
- next, if the hemisphere failed the criterias above, the Build-Up vertex was chosen, if successful, and had a decay length error smaller than 200 μm,
- lastly, if the hemisphere failed the criterias above, the Semileptonic algorithm was used, if the algorithm was successful, and the angle between the lepton and D candidate satisfied $0.8 < |\cos\theta_{ld}| < 0.99$ and if the decay length error was smaller than 1 mm.

Every hemisphere which fails all of the above criterias was excluded from the fit. Figure 6.1 illustrates the decay length resolution based on the above mentioned choices compared to the standard secondary vertex fit described in Section 4.2.9. Clearly visible is the improved resolution and the vanishing of the forward bias. The efficiency of this combination of different algorithms is about 56% for b events which is more than for every single algorithm alone. For the analysis only reconstructed proper lifetimes smaller than 10 ps were accepted to avoid the worst effects of spurious very long reconstructed lifetimes while still accepting the vast majority of the data available. The number of selected hemispheres after the cuts described in this Section can be found in Table 6.4.

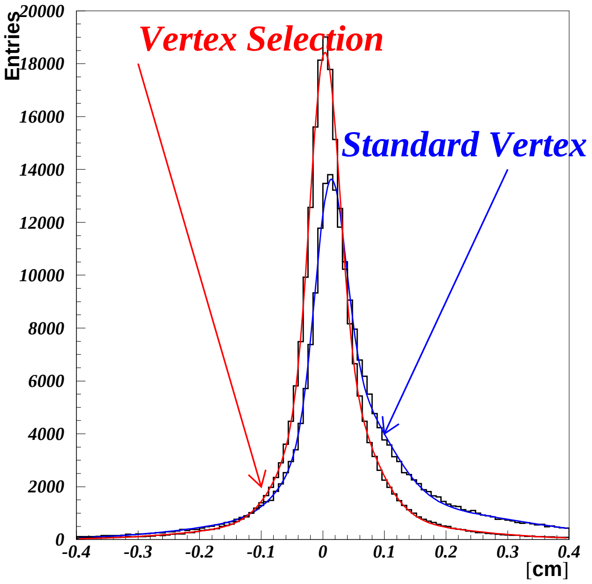

Figure 6.1: *The decay length resolution for the standard vertex algorithm of* BSAURUS *compared to the selected vertices used in the analysis is illustrated. The rather large forward bias of the standard fit is nearly gone and the resolution is improved. Three Gaussians were fitted to the distribution with:*

$$\begin{array}{ll} \bar{l}_1 = 32\,\mu m & \sigma_1 = 234\,\mu m \\ \bar{l}_2 = 118\,\mu m & \sigma_2 = 587\,\mu m \\ \bar{l}_3 = 235\,\mu m & \sigma_3 = 1670\,\mu m \end{array}$$

6.4 Monte Carlo Weighting

The first stage of the fitting procedure involved weighting quantities in the Monte Carlo to agree with recent experimental measurements. Weights were constructed to account for the following effects:

- The current world average measurements of B_s and b-baryon lifetimes (see Table 6.3). The numbers are taken from [109].
- The current measurements of the B species production fractions (see Table 6.3). The numbers are taken from [108].
- The b fragmentation function. The mean value $\langle x \rangle = 0.708$ (x is the fraction of the b-hadron energy with respect to the beam energy) of the function and the shape in the default Monte Carlo was weighted to agree with the mean value $\langle x \rangle = 0.7153 \pm 0.0052$ of the function and the shape obtained

b-hadron species	lifetime [ps]	error [ps]	fraction	error	corr. with f_{B_s}	corr. with f_{b-bary}
B_s	1.464	0.057	0.097	0.011	1	+0.034
b-baryons	1.208	0.051	0.104	0.017	+0.034	1
B^0 or B^+	this analysis		0.399	0.010	-0.577	-0.836

Table 6.3: *Results of recent experimental measurements and the world averages of the B species production fractions and the lifetimes of B_s mesons and b-baryons.*

from a recent DELPHI analysis from the Karlsruhe b physics group using an unfolding algorithm (for further details see [110]).

- Hemisphere quality. A hemisphere *quality flag* for the whole hemisphere was constructed which was correlated to e.g. the number of secondary interactions, tracks containing detector hits that could equally well fit to other tracks in the vertex detector (so-called ambiguities) or the number of tracks in the hemisphere which are excluded by the standard selection criteria of Section 4.2.2 etc. This flag is the same already mentioned in Section 4.2.9. The hemisphere quality weight was constructed as a function of hemisphere track multiplicity ensuring that the overall track multiplicity is essentially invariant under the application of the weight.

In the case of the reweighting of the production fractions, the weight is just the ratio of the measured quantities to the Monte Carlos fractions stated in Table 6.1. For the reweight of the lifetimes, the weight depends on the created true proper lifetime t^{true} via $w_i = 1.6/\tau_{wa}\exp(t_i^{true}(\tau_{wa} - 1.6)/1.6\tau_{wa})$, where τ_{wa} denotes the world average mean lifetime from Table 6.3. The quality weights are constructed by the ratio of the normalized data distribution and the normalized Monte Carlo distribution. This was done for each number of tracks in the hemisphere separately. Figure 6.2 shows the ratio of data and Monte Carlo for the distribution of the output of the hemisphere flavor Neural Network described in Section 4.2.11 before and after the application of the above described weights as an example. The lifetimes of B^0 and B^+ have been reweighted in this plot according to their world averages (1.546 ps and 1.647 ps respectively[1]). The much improved agreement between data and simulation after the weight is applied illustrates the need for such a correction.

---

[1] The numbers are taken from [109].

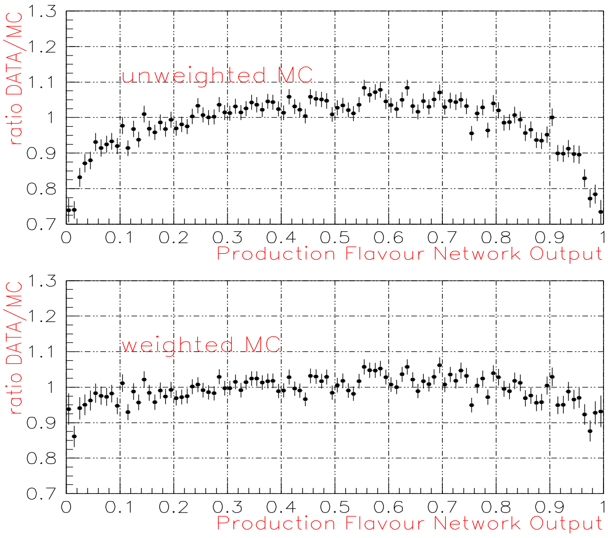

Figure 6.2: *The distributions show the ratio of data/simulation for the output of the production flavor network before (upper plot) and after (lower plot) applying the weighting procedure described in the text.*

6.5 Selection of B^+ and B^0 Enhanced Samples

The enrichment of B^+ and B^0 was part of a general attempt, implemented in BSAURUS, to provide a probability for an event hemisphere to contain a b-hadron of a particular type. The result was the Neural Network described in Section 4.2.12. The upper plots of Figure 6.3 show the weighted output of the simulation of the B^+ and B^0 output nodes of the neural net for the different b species overlaid with the data for 1994. The lower plots trace the change in purity per bin for the different b species as a function of the network output value at the B^+ and B^0 output nodes respectively.

For the analysis the neural network outputs were then cut at > 0.52 and > 0.6 to obtain enhanced samples in B^+ and B^0 respectively. These cut values correspond to a final purity in both B^+ and B^0 of $\sim 70\%$ according to the Monte Carlo. In addition, to ensure that the correlation between the two lifetimes was kept to a minimum for the final result, it was demanded that the two samples were statistically independent. This was achieved, e.g. for the case of selecting a B^+ hemisphere, by first requiring that the B^+ output node value passed a cut at 0.52. If however for this hemisphere, the B^0 output node value was also larger than the selection cut at 0.6, the hemisphere went into the B^+ sample only if the B^+ output node was also larger than the B^0 output node value. If this was not the case, the hemisphere entered into the B^0 sample. The number of events after these final selection cuts in the two enhanced samples are listed in the last two rows of Table 6.4. The final selection corresponds to a selection efficiency, with respect to the multihadronic selection sample (first row of Table 6.2) of about 13.0%(4.4%) in 1994 and 13.0%(4.4%) in 1995 for the $B^+(B^0)$ enhanced sample respectively. The overall b purity in the two samples is about 98%.

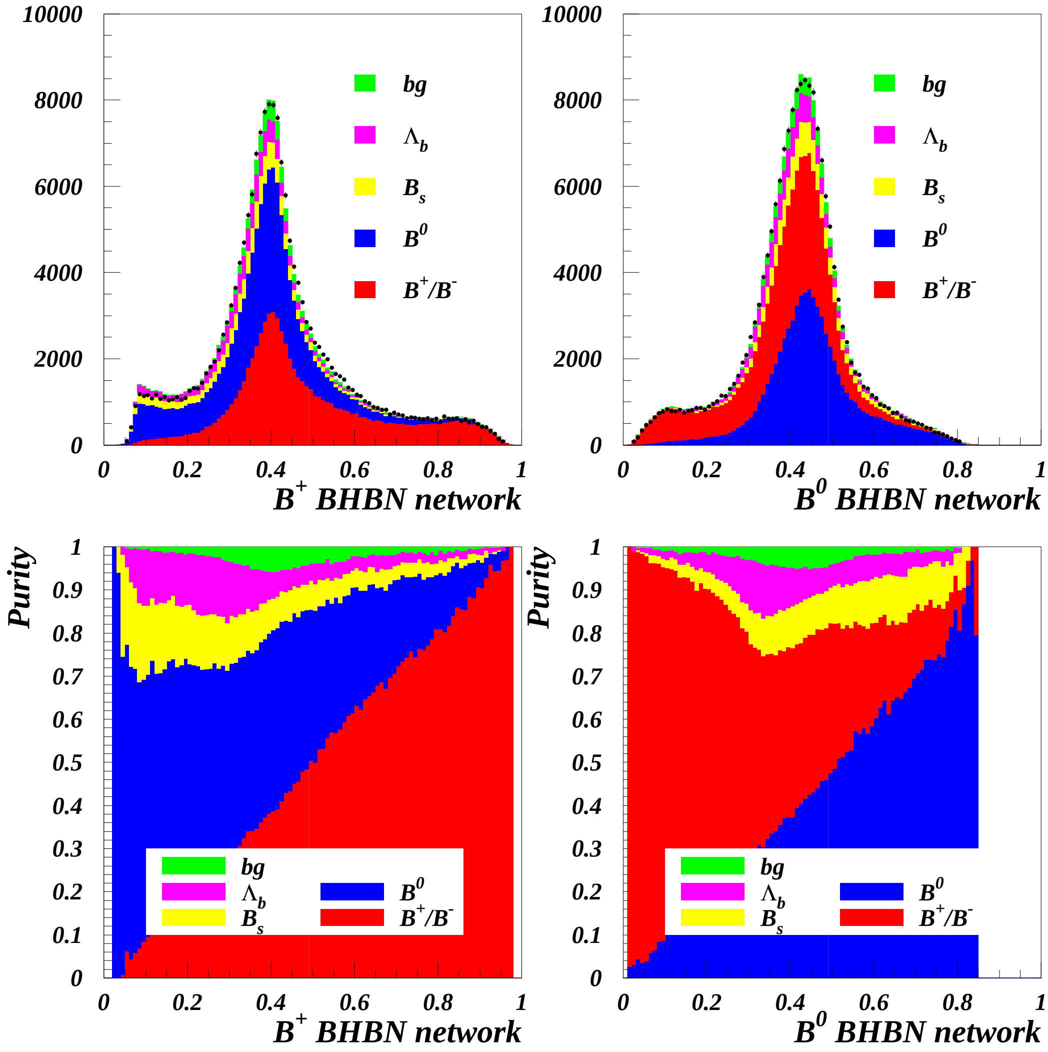

Figure 6.3: *The upper plots show the output of the* B^+ *and* B^0 *output nodes of the Both Hemispheres B-Species Identification Network (BHBN) in the 1994 data. Overlaid is the b-hadron composition as seen in the Monte Carlo and the lower plots trace the change in purity per bin. The background labeled 'bg' consists of light and charm quark events. The lower plots show the purity attained by making sequential cuts on the* B^+ *and* B^0 *outputs.*

Selection	Number of hemispheres					
	1994 data	1995 data	1994 $q\bar{q}$	1994 $b\bar{b}$	1995 $q\bar{q}$	1995 $b\bar{b}$
Vertex	180010	86796	432962	1098594	160773	292072
$t^{rec} < 10$ ps	179635	86631	432058	1096451	160463	291521
Neural Net enhancement	46878	22432	120851	313156	45545	84321
B^+ sample	36088	17367	90455	234101	33889	62645
B^0 sample	10790	5065	30396	79055	11656	21676

Table 6.4: *The number of hemispheres selected after various hemisphere cuts. The first row gives the number of hemispheres surviving the vertex selection criteria described in Section 6.3. In the second the upper decay time cut is made. The third row shows the number of hemispheres selected in the final selection cut and the last two show the division into the two enhanced samples.*

6.6 The B^+ and B^0 Lifetime Fit

The B^+ and B^0 lifetimes were extracted by a simultaneous fit to proper lifetime distributions reconstructed in the B^+ and B^0 enhanced samples, using a binned χ^2 method.

Nominally 100 bins were chosen but the exact binning was determined by the requirement that at least 10 entries be present in all bins of the data distribution to have a minimum of statistics in each bin.

To avoid the need to generate many separate Monte Carlo samples of different B lifetimes, weighting factors were formed, in a similar way than for the lifetime reweighting described in Section 6.4, for each lifetime measurement from the ratio of exponential decay probability functions. Specifically, the weight,

$$w_i = \frac{\tau_{old}}{\tau_{new}} \exp\left(\frac{t_i(\tau_{new} - \tau_{old})}{\tau_{old}\tau_{new}}\right) \quad , \tag{6.2}$$

for measurement i and true B lifetime t_i, effectively transforms the Monte Carlo lifetimes generated with a mean lifetime τ_{old} to be distributed with a new mean value of τ_{new}. The χ^2 function given below was then minimized with respect to the B^+ and B^0 lifetimes, which were the two free parameters in the fit.

$$\chi^2 = \sum_{\substack{B^+,B^0 \\ \text{sample}}} \left[\sum_{i=1}^{n_{bins}} \frac{(W_i^{MC} - N_i^{data})^2}{(\sigma_i^{MC})^2 + (\sigma_i^{data})^2}\right] \quad . \tag{6.3}$$

Here the first sum is over the two enhanced samples and the second sum is over the number of bins. N_i^{data} is the number of data entries in bin i and W_i^{MC} is the corresponding sum of weights in the Monte Carlo sample. These weights contain the weights already described in Section 6.4 as well as the weight from Equation 6.2. In addition a weight w_{lq} is constructed to account for the fact that a huge amount of $b\bar{b}$ only Monte Carlo hemispheres are used. The light and charm quark events are therefore weighted by

$$w_{lq} = \frac{N_b^{q\bar{q}} + N_b^{b\bar{b}}}{N_b^{q\bar{q}}} \tag{6.4}$$

where $N_b^{q\bar{q}}(N_b^{b\bar{b}})$ denotes the number of b quark events in the Monte Carlo $q\bar{q}(b\bar{b})$ sample respectively. This weight ensures the fraction in the distribution of light and charm quark events to be constant. The normalization of the two Monte Carlo histograms was also fixed by a weight w_s for each sample s which was equal to all hemispheres in one enhanced sample:

$$w_s = \frac{\sum_{i=1}^{n_{bins}} N_i^{data}}{\sum_{i=1}^{n_{bins}} W_i^{MC}} \tag{6.5}$$

In addition, a two parameter fit was made to the same samples to extract the ratio τ_{B^+}/τ_{B^0} of the two lifetimes separately. The two free parameters were the ratio and the arithmetic mean $(\tau_{B^+} + \tau_{B^0})/2$ of the two lifetimes.

Above described procedure has been applied to the 1994 and 1995 samples separately. The result was combined afterwards taking into account all systematic uncertainties which will be described in Section 6.8.

6.7 Working Point and Results of the Fit

The selection conditions imposed on the data samples used for the lifetime fit which were described in the previous Sections, were motivated by the wish to minimize the total error on the final results. The inexact detector resolution simulation and modeling of light and charm quark events required a relatively high b purity. The enhancement of B^+ and B^0 described in Section 6.5 requires a choice of cuts. In addition the b-hadron species enrichment Neural Network in the form described in Section 4.2.12 turned out to be problematic for the B^0 lifetime extraction which made it necessary to apply a modification to it. For small decay times the lifetime distribution of the simulation was not in good agreement with the data. Therefore a so-called *acceptance correction* was constructed. In this Section the reason for all the choices for the cuts and modifications are given. In addition the final fit results are given at the end of this Section.

6.7.1 b-Tagging Cut

The choice of the b-tagging cut at 0.3 on the combined b-tagging variable described in Section 4.1 was based on the observation of the behavior of the statistical error for varying this cut. In Figure 6.5 the top plot shows the χ^2 probability of the fit as a function of the total b purity of the two samples. The middle plot shows the b purity as a function of the cut for the two samples separate and for the two samples combined. The bottom plot shows the sum of the statistical error of the B^+ and the B^0 lifetime fit as a function of the b purity. Below 95% the error increases due to increasing pollution by not well enough modeled light and charm quark contribution. Above 99% the statistical error increases dramatically. So the cut at 0.3 which corresponds to a b purity of about 98% lies in the stable valley of the bottom plot of Figure 6.5. The χ^2 probability is quite stable and above 10% in the full region tested. In Figure 6.4 a scan over various b-tagging cuts resulting in different b purities shows a stable behavior for a relatively wide range of cuts. Only in the region were the b purity is very large, the scan becomes unstable but still lies within one statistical standard deviation according to the working point.

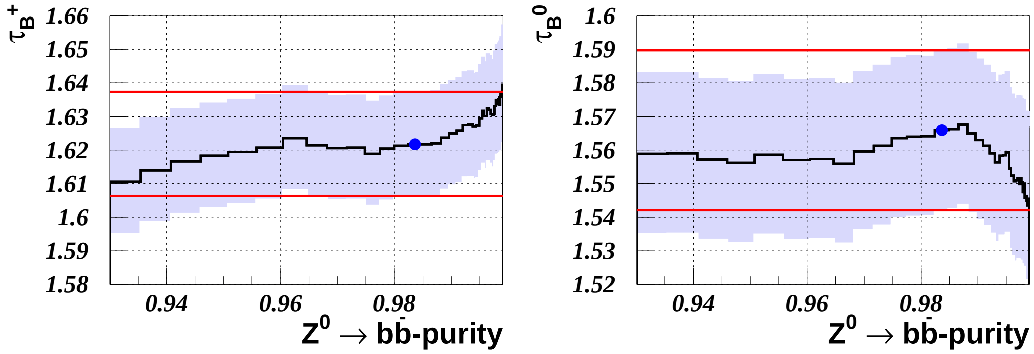

Figure 6.4: *The plots show the result of the fit of 1994 data as a function of the b purity for the B^+ lifetime fit (left) and the B^0 lifetime fit (right). The gray band indicates the one statistical standard deviation band. The lines indicate one σ of the working point (dot). The scan shows a quite stable behavior in a relatively large purity range.*

6.7.2 Neural Network Modification and Cuts

During the development of the fitting procedure various tests were made to show the stability of the result by varying key cuts. One of those tests was to change

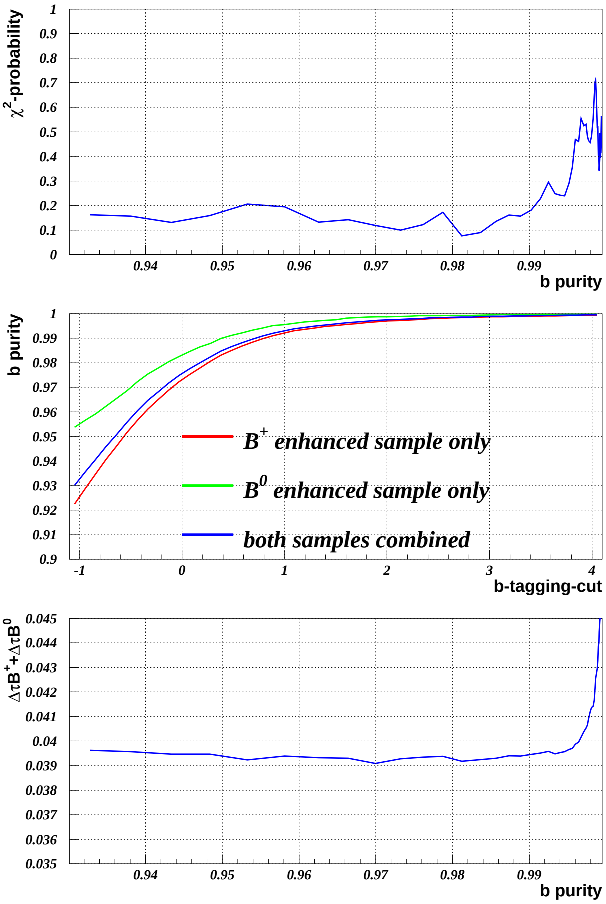

Figure 6.5: *The top plot shows the χ^2 probability for the lifetime fit as a function of various b purities. The bottom plot shows the sum of the statistical errors of the two lifetimes. The middle plot shows the b purity as a function of the b-tagging cut for the two enhanced samples separate and for both samples combined*

the fit range while basing the normalization weight from Equation 6.5 still on the full range, because no proper fit would be possible without this information. As one can see from Figure 6.6, the B^+ lifetime shows a stable behavior, whereas the B^0 lifetime shows a very significant systematic effect while increasing the lower bound or decreasing the upper bound. In the overlap region however the mean value of the left and the right plot stays stable. The reason for seeing an instability for B^0 but not for B^+ is the fact that B^0 oscillate into $\bar{B}^0$ and vice versa. In the enhancement network flavor information from the opposite hemisphere was used (see Section 4.2.12), so that the neural network learned to use this information to identify B^0 by the fact that they have oscillated. Therefore a cut on the neural network to enhance B^0 enhances more oscillated B^0 than non-oscillated B^0. The oscillation probability is time dependent and therefore the proper decay time distribution of an enhanced sample is affected. Due to inexact modeling of the Monte Carlo simulation this effect is a source of instability.

To counter this effect and to increase the stability the neural network was modified. All flavor information from the opposite hemisphere and of course explicit oscillation information was removed from the inputs. In the case of

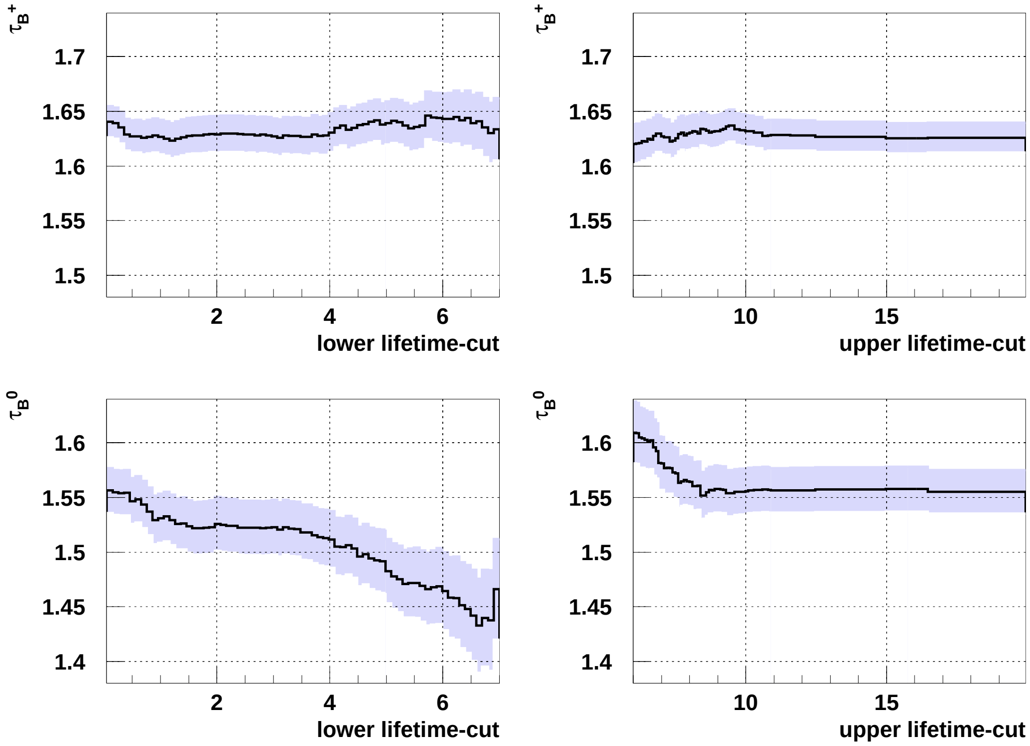

Figure 6.6: *The upper plots show the stability of the extracted B^+ lifetime for 1994 as a function of the start point (left) and end point (right) of the fit. The lower plots show the same distributions for the B^0 lifetime. The grey band marks the statistical error. While the B^+ behavior is quite stable the B^0 stability distribution shows a significant systematic effect while the upper fit bound decreases or the lower bound increases.*

flavor information, the flavor net output of the opposite side was set to 0.5 which means no information. The performance of the neural network did not suffer as much as one might expect because the most important ingredients, namely the outputs of the single hemisphere neural network 4.2.10 were kept. In Figure 6.7 the behavior using the modified neural network is illustrated. The stability of the B^0 lifetime is still not perfect but clearly improved compared to Figure 6.6.

After fixing the neural network itself an appropriate working point has to be found. The cuts on the neural network to enhance the two samples were varied. As one can see from Figure 6.8 the fit results were quite stable with respect to a cut variation in the B^+ output node. Also the increase in the error (see Figure 6.9) was smaller for the B^+ cut variation than for B^0 cut variation. For the fit it does not make sense to go below 50% purity which means no real enhancement.

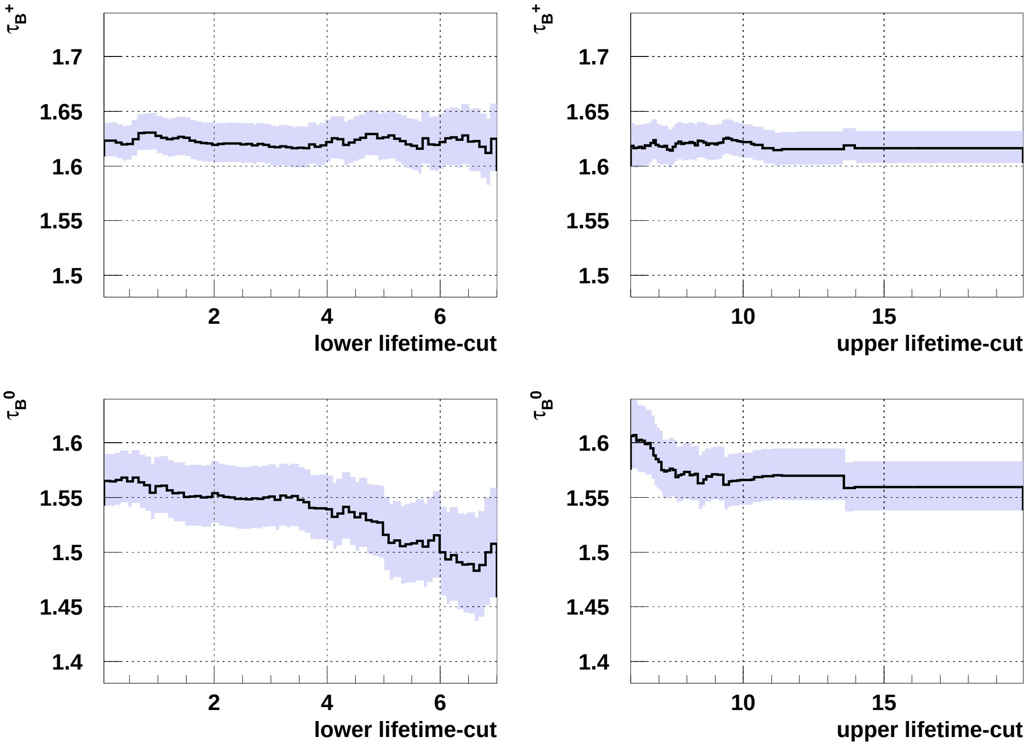

Figure 6.7: *The plots shown are the same as in Figure 6.6, but with the modified neural network used. The stability of the B^0 lifetime is still not perfect but clearly improved.*

Increasing the purity by more than 80% means increasing instability and statistical errors. Therefore the working point was chosen at 70% purity for both B^0 and B^+, which lies in a stable region and where the result does not yet suffer severely from low statistics.

6.7.3 Acceptance Correction

The efficiency for reconstructing proper lifetimes in the region below about 1.0-1.5 ps exhibits a miss-match between data and simulation. The effect is illustrated in Figure 6.10 which shows the ratio of data to simulation in this region for the case of 1994 and 1995 data separately. This observation is not surprising since this region is very sensitive to details of reconstruction resolution and the modeling of events which contain no intrinsic lifetime information such as u, d, s events and the reconstruction of spurious vertices. From Figure 6.10 it is also clear that the effect is somewhat larger in 1994 data than in 1995. The effect was corrected for

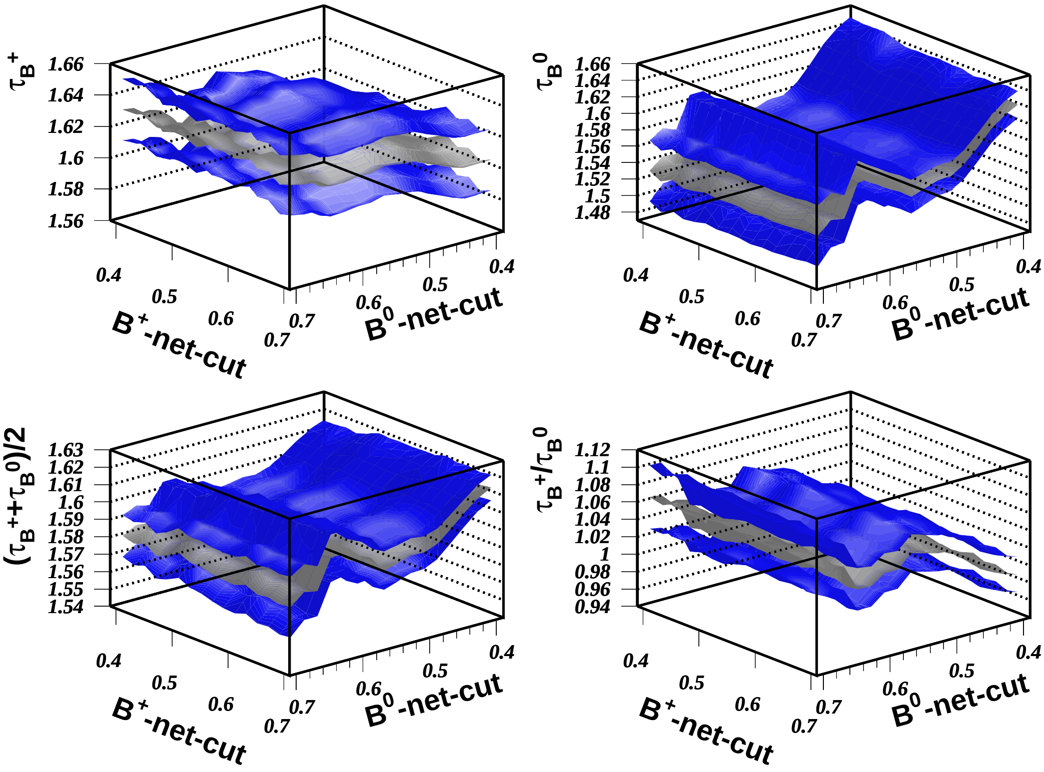

Figure 6.8: *The plots show the stability of the results with respect to the enhancement cuts. The black surfaces illustrate the result plus and minus one statistical standard deviation. The plots show the stability for the* B^+ *lifetime (top left), the* B^0 *lifetime (top right), the lifetime ratio (bottom right) and the arithmetic mean of the two lifetimes (bottom left). Varying the cut on the* B^+ *enhancement network results in a stable behavior whereas the* B^0 *lifetime is only stable in a small region around the working point.*

by fitting the shape to a parabola shown in Figure 6.10. This function was used as an additional acceptance correction weight applied on all Monte Carlo events with reconstructed decay times smaller than 2 ps.

6.7.4 Further Crosschecks

A first check of the fitting procedure was performed by using half of the available $q\bar{q}$ Monte Carlo statistics as 'data', to see if the fit finds the correct answer. For this purpose the 'data' was modified to have different lifetimes for B^+ and B^0. This has been achieved by rejecting events with a certain probability which is based on the true decay time of the meson. In this way the B^+ lifetime was

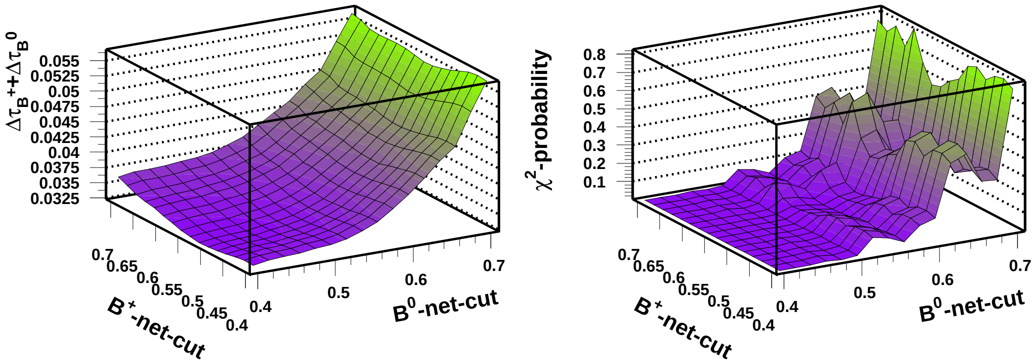

Figure 6.9: *The left plot shows the sum of the statistical error of the B^+ and B^0 lifetime fit as a function of the enhancement cuts. The right plot shows the according χ^2 probability.*

modified to be 1.5 ps whereas the B^0 lifetime was modified to be 1.7 ps. The result of this test was $\tau_{B^0} = 1.7202 \pm 0.0322$ ps and $\tau_{B^+} = 1.4928 \pm 0.0193$ ps which was in good agreement with the right answer.

A further crosscheck on the stability and robustness of the results was made by repeating the fits for one data set using the simulation sample compatible with another data set e.g. fitting 1994 data using 1995 simulation. This provides a rough check that aspects of the detector and physics modeling effects are well under control and gives information about the sensitivity to the details of the acceptance correction. It was found that all fit results (for τ_B^+, τ_B^0 and their ratio) for both cases (1994 data using 1995 simulation and 1995 data using 1994 simulation) changed by amounts that were within the systematic error assigned for detector effects (see Section 6.8).

Another check concerned the rather bad agreement between data and simulation of the BDNet (see Chapter 4 Figure 4.9). Since the Strip-Down algorithm is the only algorithm used which has an explicit cut on the output of the BDNet the analysis was run leaving out completely the Strip-Down algorithm and replacing it with the results of one of the other 3 algorithms that are employed. This gave a shift of only -0.0017 ps for the B^+ lifetime and -0.0063 ps for the B^0 lifetime. In any case the reasons for the discrepancy between data and Monte Carlo in the BDNet distributions can in principle be due to all aspects of B and, in particular, $B \to D$ modeling. It is known, that at least some part of the discrepancy comes from the inadequate modeling of ($B \to D\bar{D}X$) in the Monte Carlo and so an explicit systematic error have been assigned for this effect (see Section 6.8).

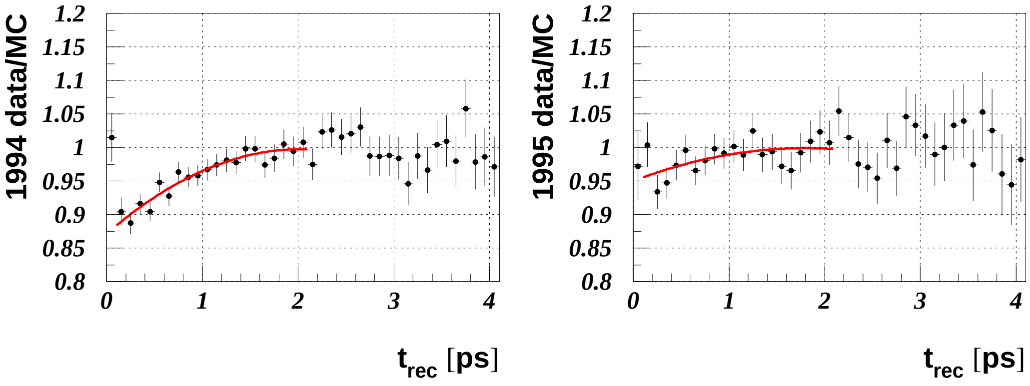

Figure 6.10: *The left plot shows the ratio of the reconstructed lifetime of data and Monte Carlo for 1994. The right plot shows the same for 1995. The superimposed functions are the fitted parabolas.*

6.7.5 Extraction of the Results

With all these modifications and corrections applied, the fit was done, after imposing the working point conditions and following the procedure described in Section 6.6. The result of the fit is illustrated in Figure 6.11 which shows the data distributions together with the best fit Monte Carlo distributions. Imposed are also the contributions from different sources, with purities and efficiencies as stated in Table 6.5. The results of this fit is stated in Table 6.6 together with the correlation and the χ^2 probability of the fit for 1994 and 1995.

An approach was made to improve the modeling of the detector response. Therefore an attempt to better match the track impact parameter and error (with respect to the primary vertex) between simulation and data according to the prescription detailed in [111] was made. Another correction was the hemisphere quality weight described in Section 6.4. The effect of those corrections was studied by switching them on and off. Since in general, knowledge of detector modeling uncertainties are not at the same level of certainty as e.g. the knowledge that b production fractions in our Monte Carlo generator differ with the world average, the result obtained from the fit by one fixed set of corrections is not more accurate than a result from any other set of corrections. We have taken the following approach to assigning systematic values for these effects: All four combinations of switching the above two corrections on and off in the analysis were made and the fitted lifetimes of the four possibilities recorded. The central results were then chosen to be the arithmetic means of these four combinations. For the statistical uncertainty the error of the default fit using the hemisphere quality correction but not the resolution matching was used, since they did not vary much.

The results of these four combinations are stated in Table 6.7 together with

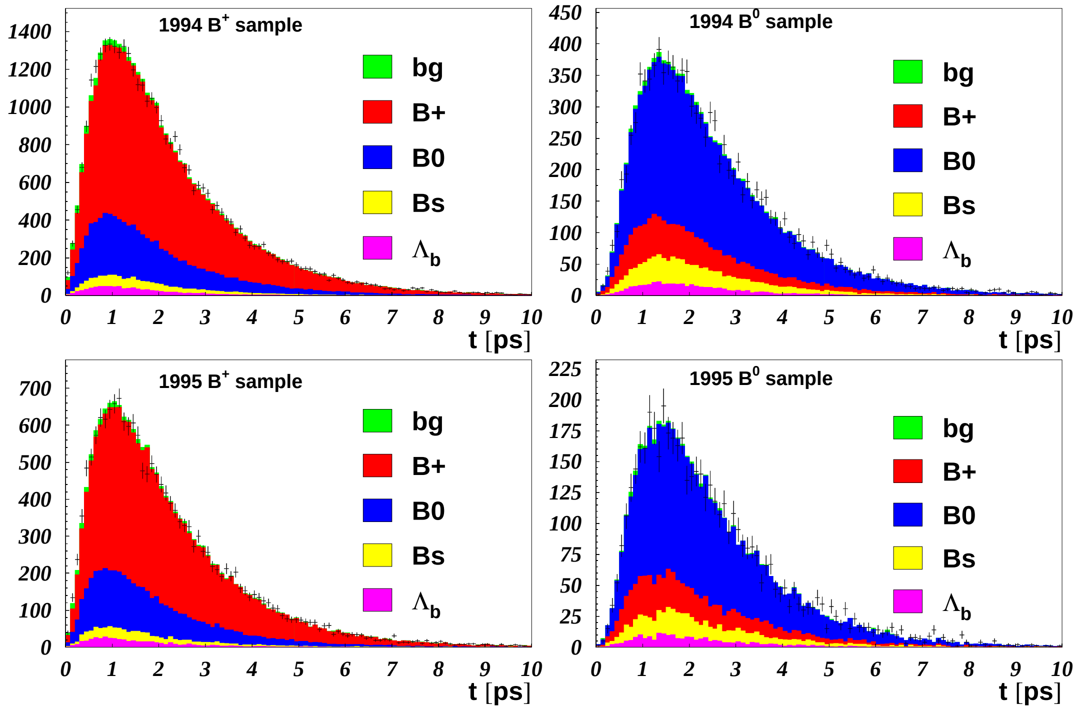

Figure 6.11: *The result of the fit in the* B^+ *(left side) and* B^0 *(right side) samples compared to the data for 1994 (top) and 1995 (bottom). The b-hadron composition of the samples is also shown. Here, bg refers to the background from non-*$b\bar{b}$ Z^0 *decays which is added on top of the plot and is hardly visible.*

the mean values.

The standard Monte Carlo data set used was generated with a wrong-sign[2] charm rate of only 11% and contained no wrong-sign D^0 or D^+ production. It is now known (see e.g. [112]) that there is an additional wrong-sign contribution, of a comparable size, due to the production of D^0 and D^+ at the W or upper vertex. To estimate the effect of this omission, the wrong-sign D_s branching ratio was enhanced to 20% with the effect of increasing the reconstructed lifetime in the Monte Carlo (or equivalently decreasing the fitted lifetime). The final fit central values were corrected downwards by the total observed shift in the fitted lifetime and in the case of the ratio upwards by the total observed shift, because the ratio behaves different.

The corrected central values which are the final values are stated in Table 6.7

---

[2] Wrong-sign charm decays are b decays, where a charm quark comes from the W-decay and therefore has an opposite sign to the charm quark from the $b \to c$ transition.

	B^+ sample 1994(1995) in %	B^0 sample 1994(1995) in %
B^+ fraction	70.6(71.0)	16.6(16.9)
B^0 fraction	21.9(21.7)	69.6(69.2)
B_s fraction	2.9(2.9)	7.8(8.2)
b-baryon fraction	2.7(2.7)	4.8(4.7)
b fraction	98.1(98.3)	98.8(99.0)
uds fraction	0.3(0.2)	0.1(0.1)
c fraction	1.6(1.4)	1.1(0.9)
	both samples combined	
b fraction	98.2(98.5)	
uds fraction	0.3(0.2)	
c fraction	1.5(1.3)	
b efficiency	10.2(10.2)	
uds efficiency	0.01(0.01)	
c efficiency	0.2(0.2)	

Table 6.5: *The B^+ and B^0 sample composition and efficiencies*

Observed quantity	Result		χ^2/n.d.f.		Corr. with τ_{B^0}
	1994	1995	1994	1995	
τ_{B^+} [ps]	1.6218 ± 0.0155	1.6231 ± 0.0230	207/179	160/161	-0.54
τ_{B^0} [ps]	1.5659 ± 0.0238	1.4942 ± 0.0358			1.00
τ_{B^+}/τ_{B^0}	1.0375 ± 0.0226	1.0865 ± 0.0368			

Table 6.6: *The results of the lifetime fits in the 1994 and 1995 data samples where the errors quoted are statistical only.*

Combination of the Results

The results obtained from 1994 and 1995 represent two independent measurements of the same quantities. In order to combine them to one single result, the following χ^2 function had to be minimized with respect to the combined result for each of the three quantities analyzed:

$$\chi^2 = \sum_{i,j} (\bar{\tau} - \tau_i) \cdot V_{ij}^{-1} \cdot (\bar{\tau} - \tau_j) \tag{6.6}$$

where V denotes the error matrix, $\bar{\tau}$ the combined result and τ_i the result i where the sum of Equation 6.6 goes over the number of measurements, here two. V is a 2×2 matrix which consists of three contributions, the statistical uncertainties σ_i, which are uncorrelated, systematic uncertainties s_i^{corr} based on physics which

Combination of correction		observed quantity					
hem. qual. weight	resol. match.	τ_{B^+} [ps]		τ_{B^0} [ps]		τ_{B^+}/τ_{B^0}	
		1994	1995	1994	1995	1994	1995
on	off	1.6218	1.6231	1.5659	1.4942	1.0357	1.0865
off	off	1.6312	1.6308	1.5899	1.5198	1.0260	1.0731
on	on	1.6420	1.6261	1.5739	1.5084	1.0433	1.0781
off	on	1.6504	1.6339	1.5996	1.5342	1.0319	1.0651
mean values		1.6364	1.6285	1.5823	1.5142	1.0342	1.0757
ws-charm 11% → 20%		1.6139	1.6144	1.5527	1.4819	1.0396	1.0894
final corr. values		**1.6285**	**1.6198**	**1.5691**	**1.5019**	**1.0381**	**1.0786**

Table 6.7: *The results of the lifetime fits in the 1994 and 1995 data samples for four different combinations of resolution corrections and for doubling the wrong-sign charm contribution using the combination of the first row. The number at the bottom which is the result of the analysis is the central value from the four numbers above shifted by half of the change in the lifetime fit according to the wrong-sign charm doubling.*

are treated conservatively as 100% correlated and other systematics s_i^{uncorr} which are treated as uncorrelated. The listing of systematics and exact division into correlated and uncorrelated error will be described in the next Section. The error matrix V then becomes:

$$V_{ij} = \sigma_i \sigma_j \delta_{ij} + \sum_{\substack{uncorrelated\\errors}} s_i^{uncorr} s_j^{uncorr} \delta_{ij} + \sum_{\substack{correlated\\errors}} s_i^{corr} s_j^{corr} \tag{6.7}$$

Minimizing the χ^2 function from Equation 6.6 leads to

$$\bar{\tau} = \frac{\sum_{i,j} V_{ij}^{-1} \cdot \tau_i}{\sum_{i,j} V_{ij}^{-1}} \tag{6.8}$$

with a total error of

$$\bar{\sigma} = \left[\sum_{i,j} V_{ij}^{-1} \right]^{-1/2} \tag{6.9}$$

To get an estimate of the statistical and systematical error separate, the total statistical error was calculated similar to Equation 6.9 with a reduced error matrix and then quadratically subtracted from the total error to get the total combined systematic uncertainty. The χ^2 can be used to calculate a probability with effectively one degree of freedom which can be interpreted as how consistent the two measurements are to each other. Then the total result becomes:

$$\boxed{\begin{array}{rclll}
\tau_{\mathbf{B^+}} & = & \mathbf{1.625}\ \pm\mathbf{0.013}\ (\mathbf{stat})\ \pm\mathbf{0.017}\ (\mathbf{syst})\ \mathrm{ps} & \text{at } \mathbf{81.9}\%\ \chi^{\mathbf{2}}\ \text{prob.} \\
\tau_{\mathbf{B^0}} & = & \mathbf{1.543}\ \pm\mathbf{0.020}\ (\mathbf{stat})\ \pm\mathbf{0.033}\ (\mathbf{syst})\ \mathrm{ps} & \text{at } \mathbf{28.6}\%\ \chi^{\mathbf{2}}\ \text{prob.} \\
\frac{\tau_{\mathbf{B^+}}}{\tau_{\mathbf{B^0}}} & = & \mathbf{1.051}\ \pm\mathbf{0.019}\ (\mathbf{stat})\ \pm\mathbf{0.024}\ (\mathbf{syst}) & \text{at } \mathbf{48.0}\%\ \chi^{\mathbf{2}}\ \text{prob.}
\end{array}}$$

6.8 Systematic Uncertainties

Systematic uncertainties on the lifetime measurements come from three main sources. The first source is from the modeling of heavy flavor parameters in our Monte Carlo. Since attempts were made to model these effects using current world average values, these errors are largely irreducible. The second source comes from the analysis method itself and the choices made in determining the measurements working point and the details of the corrections mentioned in the previous Sections. The overall good level of agreement between simulation and data and the demand that the results are stable within a wide range of the working point (as already shown e.g. in Figure 6.4) means that these errors are kept to a minimum. The third source of systematic uncertainty can be generically termed *detector effects* and result from a less than perfect modeling in simulation of the response of the detector. In this Section all systematic error sources considered are explained in detail.

6.8.1 Physics Modeling

Where possible, b physics modeling uncertainties were estimated by varying central values by plus and minus one standard deviation and taking half of the observed in the fitted lifetime value as the resulting systematic uncertainty from that source. In this way the systematic error from the lifetimes of B^s mesons and b-baryons was extracted by using the values and errors from Table 6.3.

Charged b Multiplicity

The charged multiplicity and its modeling could in principle have an effect on the measured lifetime. A recent DELPHI measurement [115] resulted in a mean charged multiplicity of 4.97 ± 0.03 (stat) ± 0.06 (syst). In our Monte Carlo we have about the same mean value. Studies were a weight was constructed to shift the mean by ±0.1 showed only small impact on the lifetime measurement and therefore no systematic error was assigned.

b-Hadron Production Fractions

The b-hadron production fractions were varied one by one by a small variation Δf_i and the total systematic error from this source was then extracted by error propagation taking account of the covariance matrix V listed in Table 6.3:

$$\sigma_\tau^2 = \sum_{i,j=1}^{3} \frac{\Delta\tau}{\Delta f_i} V_{ij} \frac{\Delta\tau}{\Delta f_j} \tag{6.10}$$

D-Topological Branching Fractions

The uncertainty from D-topological branching fractions which is the fraction of D decaying into a certain number of charged tracks, was estimated from the difference in the fit result obtained when weighting according to the results from [113]. Here the full difference was taken as a systematic error.

b-Fragmentation Function

The systematic error due to uncertainty in the b-fragmentation function was set to half of the change in the fit result seen when the $\langle x \rangle$ value was varied by $\pm 1\sigma$ of the measurement [110].

$b \to$ Charm Modeling

Close attention was paid to possible systematic effects on the analysis due to the modeling of $b \to$ charm branching ratios, where current experimental knowledge is scarce. The charm content impacts on the performance of the B^+ and B^0 enhancement networks and can pull the reconstructed b vertex to longer decay lengths. The size of this pull in turn depends on whether a D^0 or D^+ was produced since τ_{D^+} is about 2.5 times larger than τ_{D^0}. Specific aspects of the Monte Carlo that were found to warrant systematic error contributions were:

(a) As mentioned in the previous Section, the wrong-sign charm contribution in our Monte Carlo is not in agreement with latest measurements. Considering this error source we decided to shift the mean fitted lifetime values according to the description in the previous Section. Additionally a systematic error was assigned by taking half of the change in the fit result seen when changing the wrong-sign charm contribution by a reasonable amount of ±3.3%.

(b) $BR(b \to D_s X)$ is currently known to, at best, ±30% [107] and was varied in the Monte Carlo by a factor of two. The full change in the fitted lifetime was then assigned as a systematic error.

(c) The Branching ratios for $\bar{B}^0 \to (D^+)D^0X$ and $\bar{B^-} \to (D^+)D^0X$ for which no precise measurements exist were adjusted in the Monte Carlo according to a fit using all currently available measurements from [107] as constraints. This fit derived the exclusive branching fractions by using the measured inclusive branching fractions measured at LEP and at b factories, measured production fractions of b-hadrons at LEP and measured probability of prompt D^{**} decay into D. The free parameters were then the production fractions of D^{**}, D^* and D at the lower vertex (right-sign decay), D^* and D at the upper vertex (wrong-sign decay) and some other quantities. The exclusive branching ratios were extracted by considerations like isospin conservation for some decays and so on. The full shift in the fitted lifetime by putting the fitted branching fractions as a weight into the analysis was taken as a systematic error.

6.8.2 Analysis Method

Start Point of the Fit

From Figures 6.6 and 6.7 one can see that the choice of end point is rather uncritical for the measurement of the B^+ lifetime. After the modification of the neural net the B^0 lifetime is also stable in a long range of the end point (see Figure 6.7) For these reasons no systematic error was assigned to a change in the end point. The start point however was treated different with respect to the acceptance correction applied which will be discussed in Section 6.8.3.

$b(c)$ Purity and Efficiency

While changing the b purity the fit result remains stable over a wide range (see Figure 6.4) and therefore does not give the occasion to assign a systematic error, studies showed that the efficiency, especially the charm efficiency is wrongly modeled in the Monte Carlo. This efficiency for the reconstruction of $b\bar{b}$ and $c\bar{c}$ events (as a function of the event b tag) has been evaluated using a double hemisphere tagging technique on the Delphi data set [114]. At the b-tag value of the working point, the results of this study suggests that while the reconstruction efficiency for $b\bar{b}$ events might be underestimated in the simulation by about $\sim 2\%$ (relative), the efficiency for $c\bar{c}$ events is $\sim 8\%$ (relative) underestimated in the simulation compared to data. To account for this possible source of error, the efficiencies were changed in the simulation according to these results and the difference seen in the fit results assigned as a systematic error. Since a large part of the discrepancy between simulation and data in the $c\bar{c}$ event reconstruction efficiency is probably due to a relatively poor modeling of D physics, this error contribution has already, in part, been accounted for by the explicit D physics systematics detailed above. Given the current level of uncertainty in this sector, the conservative approach of quoting both error contributions is preferred.

Binning Effects

Uncertainties resulting from the method itself have been checked by varying the binning used for the χ^2 formulation. The fit has been repeated using nominally 50, 75, 150, 200 and 250 bins, the result of which is illustrated in Figure 6.12. From these plots one can see no systematic effect since the variation is well within one statistical standard deviation. Therefore no systematic error was assigned. The minimum numbers of entries per bin was also varied over a wide range and no dependence was observed.

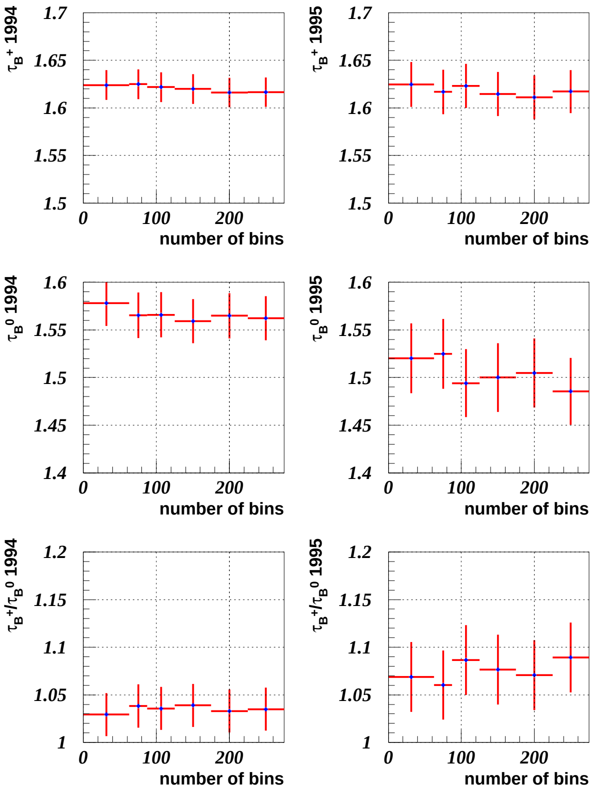

Figure 6.12: *The plots show the effect of varying the number of bins for the χ^2 formulation. The left column shows the results with systematic errors for the B^+ lifetime (top), the B^0 lifetime (middle) and their ratio (bottom) for 1994. The right column shows the same for 1995.*

B^+ and B^0 Enhancement Purities

Rather critical to the analysis is the assumption that the B^+ and B^0 purity as given by the simulation, models the situation in data well. The effect in the result of scanning over the B^+ and B^0 enhancement purities was already illustrated in

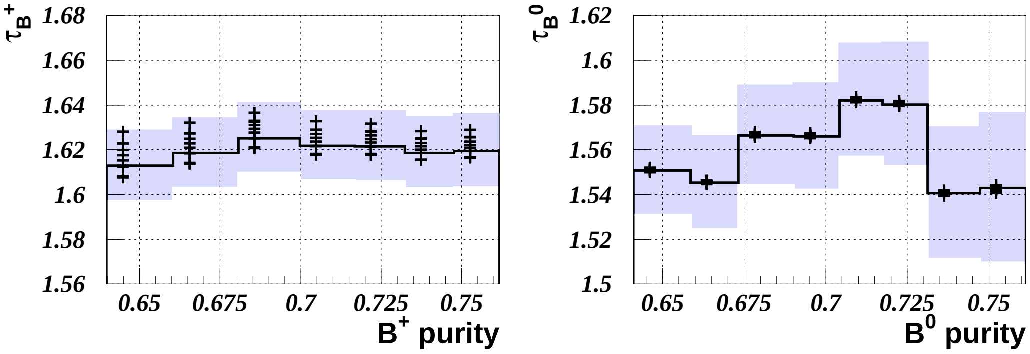

Figure 6.13: *The plots show the variation in the fitted* B^+ *lifetime (left) and* B^0 *lifetime (right) as a function of the enhancement cuts on the* B^+ *(left) and* B^0 *output node. The upper and lower shaded bands represent the statistical one standard error. The spread of the points for each bin of purity represents changing the other enhancement purity in the range* $[65\%, 75\%]$.

Figure 6.8. The same is shown in more detail in Figure 6.13. A systematic error will arise if the composition of the B^+ and B^0 simulated samples differ from the data and/or the overall shape of the B^+ and B^0 output node distributions differ. To account for the first effect, half of the maximum variation in the fitted lifetime while scanning the purity range $[65\%, 75\%]$ was assigned as an error (see Figure 6.13). A separate error contribution was also assigned to take account of any residual difference in shape between data and simulation in the net distribution. Assuming that the difference can be wholly accounted for by a change in the B^+ or B^0 composition in the sample, it was found that the maximum error made in calculating the B^+ or B^0 purities in the samples fitted was of order 2% and 4% respectively. The effect of these changes were then propagated into errors on the extracted lifetimes.

Secondary Vertex Multiplicity

The multiplicity of tracks in the reconstructed b-hadron vertex was found to be in overall good agreement between data and simulation. To account for any residual differences, mainly in the tails of the distribution (see Figure 6.14) a weight was formed from the ratio of the data and simulation distributions (Figure 6.14) and the change seen as a result of applying this weight was assigned as a systematic error.

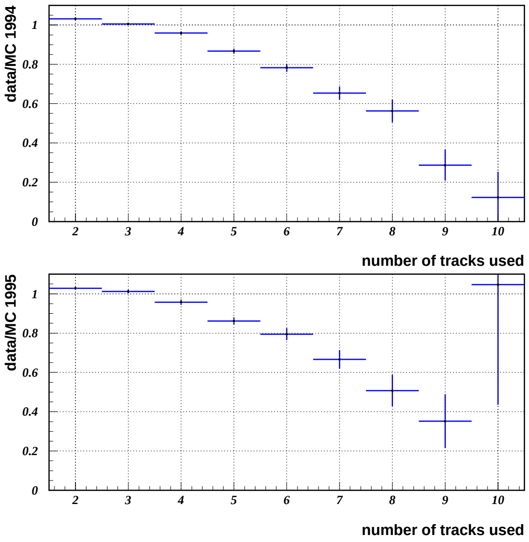

Figure 6.14: *The plots show the ratios of data and Monte Carlo distributions for 1994 (upper plot) and 1995 (lower plot) for the number of tracks used for the secondary vertex algorithm.*

6.8.3 Detector Effects

Acceptance Correction

As mentioned in Section 6.7.3, modeling the region at low proper time is a particularly complex task. That this is the case is illustrated by Figure 6.15 which shows that, in the region below about 2 ps in the B^0/B^+ samples, the purity of the different b-hadron species is rapidly changing. If the variation of the purities with proper time are not accurately modeled in the simulation one must expect to reconstruct unstable lifetime fit results as a function of the point in proper time chosen to start the fit from. Such lifetime scans are presented in Figure 6.16 up to 4 ps. The results show that, in general, the fit results are rather stable which gives confidence that the low proper time region is reasonably modeled. For the case of B^0 the acceptance correction described in Section 6.7.3 helped to reduce the effect but some residual systematic trend is still visible. To account for this, and to reflect the overall uncertainty associated with this region, a systematic error (labeled 'acceptance modeling' in Tables 6.8 and 6.9) was assigned to the B^+ and B^0 results in 1994 and 1995 equal to half of the full spread in results seen over the interval $[0.0 \text{ ps}, 2.0 \text{ ps}]$ with the acceptance correction not applied. The upper range of this interval was chosen to coincide with the point in proper time at which the purities stabilized in our samples as already shown in Figure 6.15. Of course the values to which these purities stabilize to can also be wrongly modeled in the simulation and this possibility has been systematically accounted for by the error assigned for the b-hadron production fractions and the errors associated with the cuts and shape of the enhancement neural networks.

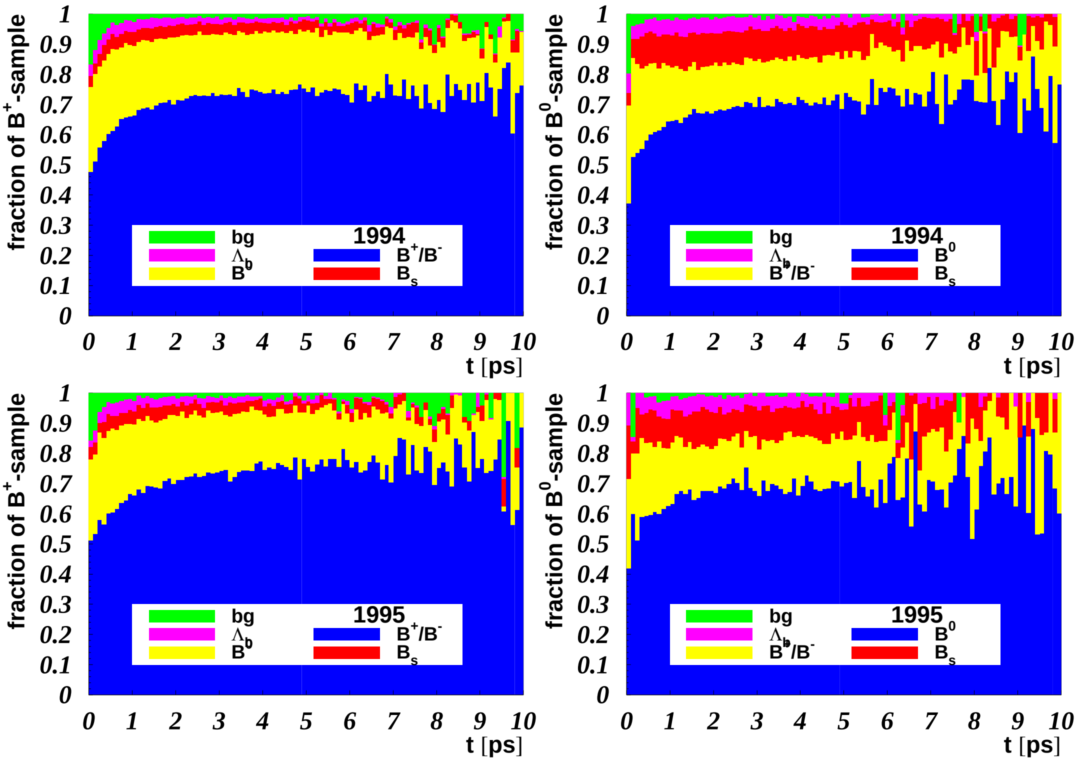

Figure 6.15: *The plots trace how the fractional composition of the samples change in bins of the reconstructed lifetime for the B^+ (left column) and B^0 (right column) samples in 1994 (top row) and in 1995 (bottom row).*

Detector Modeling

As already mentioned, in order to account for uncertainties in the simulation originating from detector response modeling, the effect was studied of switching on and off the two corrections described in Section 6.7.5, namely the resolution smearing and the quality weight. The resulting systematic error from detector response modeling was assigned to be half of the full spread of the values from the four combinations.

Tables 6.8 and 6.9 present the full systematic error breakdown for the measurements of τ_{B^+}, τ_{B^0} and their ratio in 1994 and 1995 data respectively. In general it can be seen that detector effects dominate over physics modeling errors. The extremely high $Z^0 \to b\bar{b}$ purity of the event samples used means that only modeling errors from b physics are of importance. Of these, only the wrong-sign charm contribution is sizeable due to the fact that our default Monte Carlo did not account for recent evidence for a larger branching ratio $b \to c\bar{c}s$. There are also relatively large systematic error contributions arising from the enhance-

1994 Data		τ_{B^+}	τ_{B^0}	$\frac{\tau_{B^\pm}}{\tau_{B^0}}$
Result		**1.6285**	**1.5691**	**1.0381**
Statistical Error		**0.0155**	**0.0238**	**0.0226**
Source of Systematic Error	**Range**	$\Delta\tau_{B^+}$	$\Delta\tau_{B^0}$	$\Delta\frac{\tau_{B^\pm}}{\tau_{B^0}}$
	Physics Modeling			
B_s lifetime	1.464 ± 0.057 ps	0.0007	0.0079	0.0048
b-baryon lifetime	1.208 ± 0.051 ps	0.0008	0.0028	0.0030
b-hadron prod. fractions	See text	0.0034	0.0038	0.0003
frag. function	$\langle x \rangle = 0.7153 \pm 0.0052$	0.0037	0.0020	0.0036
BR($B \to D$)	See text	0.0085	0.0069	0.0008
BR($B \to$ wrong-sign charm)	11% → 22%	0.0029	0.0048	0.0015
BR($B_s \to D_s$)	35% → 70%	0.0019	0.0075	0.0061
D^+,D^0 topo. branching ratios	[113]	0.0002	0.0100	0.0065
B meson mass	5.2789 ± 0.0018 GeV	0.0002	0.0017	0.0010
	Analysis Method			
b and c efficiency corr.	on/off	0.0076	0.0050	0.0015
$NN(B_x)$ cuts	65%-75% purity	0.0093	0.0216	0.0196
$NN(B_x)$ shape	See text	0.0008	0.0099	0.0059
Sec. vertex multiplicity	see text	0.0022	0.0042	0.0040
	Detector Effects			
Resolution and hemisphere quality	On/off	0.0143	0.0169	0.0072
Acceptance modeling	See text	0.0071	0.0140	0.0090
Total Systematic Error		**0.0227**	**0.0375**	**0.0264**

Table 6.8: *Summary of systematic uncertainties in the B^+ and B^0 lifetime results and their ratio for 1994 data. Systematic errors are assumed independent and are added in quadrature to give the total error.*

1995 Data		τ_{B^+}	τ_{B^0}	$\frac{\tau_{B^\pm}}{\tau_{B^0}}$
Result		**1.6198**	**1.5019**	**1.0786**
Statistical Error		**0.0230**	**0.0358**	**0.0368**
Source of Systematic Error	**Range**	$\Delta\tau_{B^+}$	$\Delta\tau_{B^0}$	$\Delta\frac{\tau_{B^\pm}}{\tau_{B^0}}$
	Physics Modeling			
B_s lifetime	1.464 ± 0.057 ps	0.0007	0.0083	0.0055
b-baryon lifetime	1.208 ± 0.051 ps	0.0006	0.0034	0.0021
b-hadron prod. fractions	See text	0.0038	0.0029	0.0005
frag. function	$\langle x \rangle = 0.7153 \pm 0.0052$	0.0038	0.0038	0.0052
$BR(B \to D)$	See text	0.0072	0.0124	0.0040
$BR(B \to$ wrong-sign charm$)$	11% → 22%	0.0032	0.0045	0.0011
$BR(B_s \to D_s)$	35% → 70%	0.0013	0.0077	0.0065
D^+,D^0 topo. branching ratios	[113]	0.0048	0.0142	0.0135
B meson mass	5.2789 ± 0.0018 GeV	0.0010	0.0010	0.0014
	Analysis Method			
b and c efficiency corr.	on/off	0.0048	0.0066	0.0017
$NN(B_x)$ cuts	65%-75% purity	0.0133	0.0255	0.0269
$NN(B_x)$ shape	See text	0.0006	0.0126	0.0097
Sec. vertex multiplicity	see text	0.0019	0.0040	0.0042
	Detector Effects			
Resolution and hemisphere quality	On/off	0.0054	0.0200	0.0107
Acceptance modeling	See text	0.0071	0.0067	0.0069
Total Systematic Error		**0.0200**	**0.0430**	**0.0361**

Table 6.9: *Summary of systematic uncertainties in the B^+ and B^0 lifetime results and their ratio for 1995 data. Systematic errors are assumed independent and are added in quadrature to give the total error.*

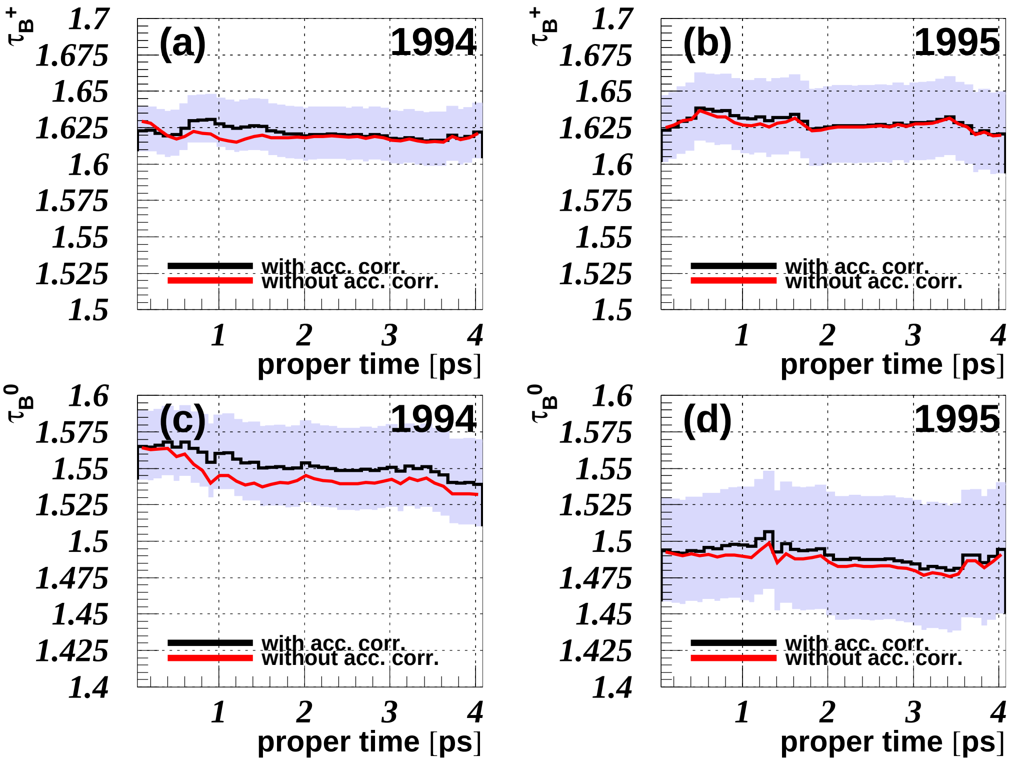

Figure 6.16: *The plots show a scan over the starting point of the lifetime fit for* B^+ *in (a) 1994 and (b) 1995 and for* B^0 *in (c) 1994 and (d) 1995. The shaded plots show the lifetime fit result with the acceptance correction applied. Overlaid is the scan without acceptance correction. The band indicates the one statistical standard deviation.*

ment neural networks (labeled as $NN(B_x)$) which were to be expected from the difficult task of modeling accurately such complex variables.

From Tables 6.8 and 6.9 one can see, that most of the physics errors were quite small with respect to the statistical error and some of them show an inconsistent size comparing 1994 with 1995. For this reason the physics modeling errors have been treated as equal in 1994 and 1995 and the weighted mean (using the statistical error) was chosen to be used for the combination. The results for 1994 and 1995 were combined afterwards, treating all systematic contributions as independent apart from the following (which were assumed to be 100% correlated):

- all physics modeling errors,
- the neural net shape error ($NN(B_x)$ shape),

- the secondary vertex multiplicity error.

6.9 Summary

An inclusive B^+ and B^0 lifetime measurement was performed using the DELPHI data set taken on the Z^0 resonance in the years 1994 and 1995.

After selecting b-hadron candidates with a purity of about 94% and reweighting of the Monte Carlo sample to correct the simulation for known discrepancies, the analysis isolated and reconstructed B^+ and B^0 meson candidate vertices and momentum with purities of about 70% using neural network techniques trained to exploit the physical properties of inclusive b-hadron decays. This treatment resulted in a total sample size of about 47k(22k) event hemispheres for 1994(1995). The reconstruction and enhancement algorithms were based on the BSAURUS package [94].

A binned χ^2 fit to the resulting DELPHI data samples fitting both lifetimes simultaneously and the ratio in a separate fit yielded the results presented in Tables 6.8 and 6.9 for the two years separately. The systematic effects and additional correction procedures of the analysis were investigated and described in detail. The largest systematic error contributions came from uncertainties in detector modeling and the details of the enhancement neural networks.

The two analysis of the two years were treated as two independent measurements and the results were combined using standard techniques. The results of this combination, which are also the final results of this analysis, are:

$$\boxed{\begin{array}{rcl} \tau_{\mathbf{B^+}} & = & \mathbf{1.625} \ \pm \mathbf{0.013}\ (\mathbf{stat})\ \pm \mathbf{0.017}\ (\mathbf{syst})\ \text{ps} \quad \text{at } \mathbf{81.9}\%\ \boldsymbol{\chi^2}\ \text{prob.} \\ \tau_{\mathbf{B^0}} & = & \mathbf{1.543} \ \pm \mathbf{0.020}\ (\mathbf{stat})\ \pm \mathbf{0.033}\ (\mathbf{syst})\ \text{ps} \quad \text{at } \mathbf{28.6}\%\ \boldsymbol{\chi^2}\ \text{prob.} \\ \dfrac{\tau_{\mathbf{B^+}}}{\tau_{\mathbf{B^0}}} & = & \mathbf{1.051} \ \pm \mathbf{0.019}\ (\mathbf{stat})\ \pm \mathbf{0.024}\ (\mathbf{syst}) \qquad \text{at } \mathbf{48.0}\%\ \boldsymbol{\chi^2}\ \text{prob.} \end{array}}$$

The total errors were therefore:

$$\sigma^{tot}_{\tau_{B^+}} = 0.022 \text{ ps}$$
$$\sigma^{tot}_{\tau_{B^0}} = 0.038 \text{ ps}$$
$$\sigma^{tot}_{\tau_{B^+}/\tau_{B^0}} = 0.031$$

Chapter 7

Measurement of the Average b-Hadron Lifetime

In this Chapter the measurement of the average b-hadron lifetime is presented. Historically the measurement of τ_b was used to determine the magnitude of the CKM matrix element $|V_{cb}|$ in an inclusive way. At this time experimental precision was not good enough to differentiate between the lifetime components in a b-hadron sample. Meanwhile precision and understanding of heavy flavor physics increased and a measurement of τ_b is less important. Nevertheless, the sophisticated reconstruction algorithms used in this analysis enabled the most precise measurement of τ_b so far.

The first part of this Chapter describes the differences between the measurement of τ_{B^+} and τ_{B^0} described in the previous Chapter and the τ_b measurement. Though in principle the same pre-selection and fitting method is used, there are some things that are treated differently.

In the second part the fitting procedure is described and the results are given.

The third part of this Chapter contains the discussion of the systematic errors considered.

Finally the fourth part gives a short summary and interpretation of the result in form of a determination of the CKM matrix element $|V_{cb}|$ which then will be compared to the world average.

7.1 Sample Selection

The event selection for the τ_b fit sample is exactly the same as for the B^+ and B^0 lifetime fit. The choice of the working point for the b-tagging cut has the same reason, to minimize the statistical error while still having a stable result. Also the

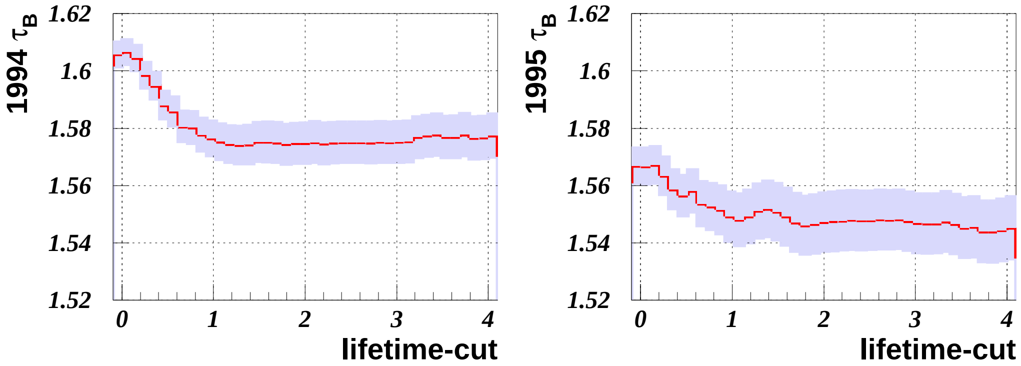

Figure 7.1: *The plots show a scan over the starting point of the average lifetime fit for 1994 (left) and 1995 (right) data. The shaded bands show the one standard deviation error. The systematic effect for the low lifetime region is the reason to start the fit at 1 ps onwards to exclude the unstable region.*

Selection	Number of hemispheres					
	1994 data	1995 data	1994 $q\bar{q}$	1994 $b\bar{b}$	1995 $q\bar{q}$	1995 $b\bar{b}$
Vertex	180010	86796	432962	1098594	239061	292072
$t^{rec} < 10$ ps	179635	86631	432058	1096451	160463	291521
$t^{rec} > 1$ ps	114317	54958	272188	701933	101835	187336

Table 7.1: *The number of hemispheres selected after various hemisphere cuts. The first row gives the number of hemispheres surviving the vertex selection criteria described in Section 6.3. In the second the upper decay time cut is made. The third row shows the number of hemispheres selected in the final selection cut using only hemispheres with a reconstructed lifetime greater than 1 ps.*

selection of the secondary vertex algorithm is the same as described in Section 6.3. The main difference, of course, is that no enhancement is required, and therefore after the cut at 10 ps on the reconstructed lifetime, the sample size remains the same. However changing the start point of the fit, in this case gives a problem (see Figure 7.1), especially for the 1994 data. Here the modeling uncertainty at low proper decay times gives a huge systematic effect on the average lifetime measurement. It turned out that this effect cannot be corrected so easily by introducing an acceptance correction as it was for the τ_{B^+} and τ_{B^0} measurement. Since this region is not well under control, it was preferred to chop away from the fit the region less than 1 ps which coincides with the point at which the b-species purities stabilize in Figure 7.2. Since the statistics in the b-hadron sample are relatively large, the reduction in the statistics resulting from making the cut at 1

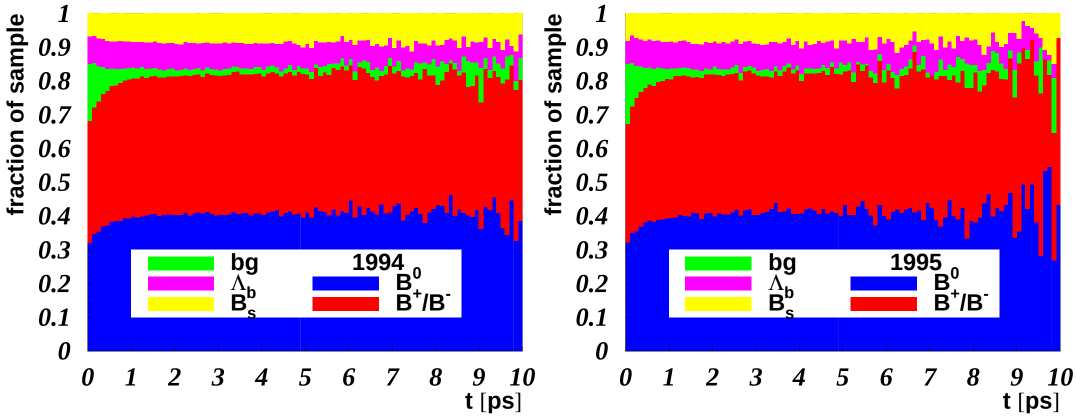

Figure 7.2: *The plots trace how the fractional composition of the samples change in bins of the reconstructed lifetime for the mean b-hadron sample in 1994 (left) and in 1995 (right).*

ps are not as important to the final precision as would be the case for the B^0 and B^+ fits. This selection resulted in the event hemisphere numbers of Table 7.1. The purities and efficiencies of Table 7.2 do also refer to the above Selection.

7.2 The Average b Lifetime Fit

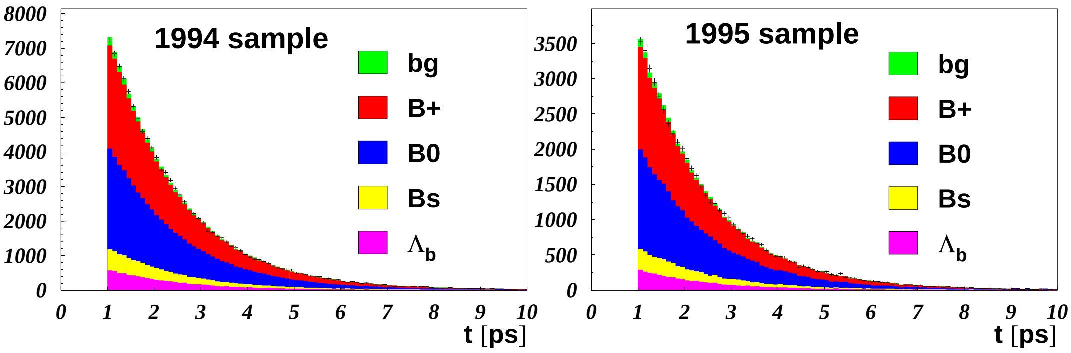

Figure 7.3: *The result of the fit in the 1994 (left) and 1995 (right) data samples compared to the data for 1994. The b-hadron composition of the samples is also shown. Here, bg refers to the background from non-$b\bar{b}$ Z^0 decays which is added on top of the plot and hardly visible.*

A one parameter fit was made to the not enhanced data sample (i.e. the sample contained the natural mix of b-hadron species like stated in Table 7.2)

	1994 sample in %	1995 sample in %
B^+ purity	42.7	42.8
B^0 purity	41.8	42.0
B_s purity	6.7	6.7
b-baryon purity	6.8	6.6
b purity	98.0	98.1
uds purity	0.3	0.3
c purity	1.7	1.6
b efficiency	22.9	22.7
uds efficiency	0.04	0.04
c efficiency	0.58	0.54

Table 7.2: *The composition and efficiencies of the τ_b fit sample*

Observed quantity	Result		χ^2/n.d.f.	
	1994	1995	1994	1995
τ_b [ps]	1.5765 ± 0.0059	1.5494 ± 0.0085	76/88	70/88

Table 7.3: *The result of the average lifetime fit in the 1994 and 1995 data samples where the errors quoted are statistical only.*

for the mean b-hadron lifetime τ_b. Therefore the same χ^2 formulation as for the τ_{B^+} and τ_{B^0} fit was chosen. The two samples of 1994 and 1995 were fitted again separately and the results combined. The result of the fit is illustrated in Figure 7.3 which shows the data distributions together with the best fit Monte Carlo distributions. Imposed are also the contributions from different sources, with purities and efficiencies as stated in Table 7.2. The results of this fit is stated in Table 7.3 together with the χ^2 probability of the fit for 1994 and 1995. To account for possible modeling uncertainties of the detector response, the same corrections as for the τ_{B^+} and τ_{B^0} fit were applied, namely the resolution smearing and the quality weight. Again the results of the four combinations of switching these combinations on and off were noted and the mean of these four numbers calculated. This mean was again corrected by the deviation to the result when the wrong-sign charm content is set to 20%. The numbers leading to the final results are stated in Table 7.4. The two measurements in the 1994 and the 1995 data sample are again combined in the same way as the results in the previous chapter, leading to the following result:

Combination of correction			
hem. qual. weight	resol. match.	τ_b [ps]	
		1994	1995
on	off	1.5765	1.5494
off	off	1.5819	1.5555
on	on	1.5783	1.5596
off	on	1.5835	1.5659
mean values		1.5801	1.5576
ws-charm 11% → 20%		1.5737	1.5464
final corr. values		**1.5773**	**1.5546**

Table 7.4: *The result of the average lifetime fit in the 1994 and 1995 data sample for four different combinations of resolution corrections and for doubling the wrong-sign charm contribution using the combination of the first row. The number at the bottom which is the result of the analysis is the central value from the four numbers above shifted by half of the change in the average lifetime fit according to the wrong-sign charm doubling.*

$$\boxed{\tau_\mathbf{b} \;=\; \mathbf{1.568}\ \pm\mathbf{0.005}\ \mathbf{(stat)}\ \pm\mathbf{0.009}\ \mathbf{(syst)}\ \text{ps} \quad \text{at } \mathbf{20.2}\%\ \chi^\mathbf{2}\ \text{prob.}}$$

7.3 Discussion of Systematic Errors

For the average lifetime the same systematic errors have been considered as in the previous Chapter for τ_{B^+} and τ_{B^0}. Without enhancement, no systematic error from the neural net cuts occurs of course. Since the fit started at 1 ps a systematic error, labeled 'acceptance modeling' in Table 7.5, is assigned for the exact positioning of this cut based on half of the change in the fit result seen when varying the cut in the range $[0.5\ ps, 1.5\ ps]$. The remaining errors can then be divided again into three main sources, physics modeling, systematics from the method used and detector modeling uncertainties. From all these errors the detector uncertainties gave the biggest contribution. From the physics systematics, which were all quite small the fragmentation function error source gave the biggest systematic effect. A full systematic error breakdown for the measurement of τ_b in 1994 and 1995 data is presented in Table 7.5. For the combination the physics modeling errors have been treated as equal in 1994 and 1995 and the weighted mean (using the statistical error) was calculated.

		$\tau_{\rm B}$ **for 1994**	$\tau_{\rm B}$ **for 1995**
Result		**1.5784**	**1.5559**
Statistical Error		**0.0059**	**0.0085**
Source of Systematic Error	**Range**	$\Delta\tau_{\rm B}$	$\Delta\tau_{\rm B}$
	Physics Modeling		
b-hadron prod. fractions	See text	0.0012	0.0010
frag. function weight	on/off	0.0039	0.0037
BR(B $\rightarrow$ wrong-sign charm)	11% $\rightarrow$ 22%	0.0010	0.0010
BR($B_s \rightarrow D_s$)	35% $\rightarrow$ 70%	0.0011	0.0015
B $\rightarrow$ D branching fractions	See text	0.0005	0.0015
D^+, D^0 topo. branching ratios	[113]	0.0011	0.0003
B meson mass	$m_B = 5.2789 \pm 0.0018$ GeV	0.0002	0.0009
	Analysis Method		
b and c efficiency correction	on/off	0.0057	0.0048
Sec. vertex multiplicity	see text	0.0019	0.0023
	Detector Effects		
Resolution and hemisphere quality	On/off	0.0035	0.0083
Acceptance modeling	See text	0.0069	0.0050
Total Systematic Error		**0.0108**	**0.0120**

Table 7.5: *Summary of systematic uncertainties in the average b-hadron lifetime for 1994 and 1995 data. Systematic errors are assumed independent and are added in quadrature to give the total systematic error.*

7.4 Summary and Interpretation

A measurement of the average mean b lifetime τ_b was performed using the DELPHI data set taken on the Z^0 peak in the years 1994 and 1995.

The selection of b-hadron candidates was essentially the same as in the τ_{B^+} and τ_{B^0} analysis leading to a b purity of about 94%. After the vertex Selection which also was the same as for the previous analysis described in Chapter 6 however no enhancement cuts were applied, to keep a natural composition of b-hadrons. The starting point of the fit was chosen to be at 1 ps to overcome the badly modeled low decay time region while still having a high amount of statistics. Since the fit was started at 1 ps also no acceptance correction was necessary. This treatment lead to a b-hadron purity of about 98% and resulted in a total sample size of about 114k(55k) event hemispheres for 1994(1995).

Again, a binned χ^2 fit to the resulting DELPHI data samples yielded the results presented in Table 7.5. The systematic effects were briefly discussed and the largest systematic error turned out to come from imperfect detector modeling.

The two analysis of the two years were treated as two independent measurements of the same quantity and the results were combined using standard techniques. The result of this combination which is also the final result of this analysis is:

$$\boxed{\tau_{\mathbf{b}} \;=\; \mathbf{1.568\ \pm 0.005\ (stat)\ \pm 0.009\ (syst)}\ \text{ps} \quad \text{at } \mathbf{20.2\%}\ \boldsymbol{\chi^2}\ \text{prob.}}$$

The total error was therefore:

$$\sigma_{\tau_b}^{tot} = 0.010 \text{ ps}$$

In Section 1.4.2 in Chapter 1 it has been described, that the average mean lifetime τ_b is related to $|V_{cb}|$. This relation is via the semileptonic charmed branching ratio $BR(b \to X_c l \bar{\nu})$ and comes with theoretical uncertainties [116]:

$$|V_{cb}| = (41.6 \pm 2.0) \times 10^{-3} \times \eta_{QED} \left(\frac{BR(b \to X_c l \bar{\nu})}{0.105} \cdot \frac{1.6 \text{ ps}}{\tau_b} \right)^{1/2} \tag{7.1}$$

where $\eta_{QED} \simeq 1.007$ is the electromagnetic radiative correction. A similar equation can be written for $|V_{ub}|$:

$$|V_{ub}| = (3.06 \pm 0.16) \times 10^{-3} \times \left(\frac{BR(b \to X_c l \bar{\nu})}{0.001} \cdot \frac{1.6 \text{ ps}}{\tau_b} \right)^{1/2} \tag{7.2}$$

These two equations can be combined to have a relation which only depends on the inclusive semileptonic branching ratio $BR(b \to Xl\bar{\nu})$:

$$\frac{|V_{cb}|^2}{(41.6 \cdot 10^{-3} \cdot \eta_{QED})^2} + \frac{|V_{ub}|^2}{(3.06 \cdot 10^{-3})^2}\frac{0.001}{0.105} = \frac{BR(b \to Xl\bar{\nu})}{0.105}\frac{1.6}{\tau_b} \tag{7.3}$$

From this follows:

$$|V_{cb}| = 41.6 \cdot 10^{-3} \cdot \eta_{QED}\sqrt{\frac{BR(b \to Xl\bar{\nu})}{0.105}\frac{1.6}{\tau_b} - \frac{|V_{ub}|^2}{(3.06 \cdot 10^{-3})^2}\frac{0.001}{0.105}} \tag{7.4}$$

Using measured values for $BR(b \to Xl\bar{\nu}) = 10.59 \pm 0.22\%$ and $|V_{ub}| = (3.6 \pm 0.7) \times 10^{-3}$ from [18] and taking the above mentioned theoretical uncertainties into account, the measured average lifetime τ_b gives a $|V_{cb}|$ value of:

$$|V_{cb}| = (42.2 \pm 2.1) \times 10^{-3} \tag{7.5}$$

where the error is extremely dominated by the theoretical uncertainty. This estimation of $|V_{cb}|$ is consistent with the current world average value which is $|V_{cb}| = (41.2 \pm 2.0) \times 10^{-3}$ in [18].

Chapter 8

Conclusion

An inclusive b hadron lifetime analysis was performed using the DELPHI data set taken on the Z^0 peak in the years 1994 and 1995.

Event hemispheres were selected, passing the multihadronic selection criteria, the b-tagging cut and the cut on the direction of the thrust axis, leading to approximately 94% purity in $Z^0 \to b\bar{b}$ events. New vertexing algorithms based on the BSAURUS package were used, which made extensive use of elaborate neural network techniques. The BSAURUS package also delivered dedicated neural networks for the identification of b-hadrons and reconstruction algorithms for the b momentum, exploiting the full capabilities of the DELPHI detector.

After all selection cuts the data sample consisted of about 420000 event hemispheres. A χ^2 fit was performed to extract the lifetimes of B^+ and B^0 mesons using samples which were about 70% pure in B^+ and B^0 mesons. The lifetimes and their ratio were measured to be:

$$\begin{aligned}
\tau_{B^+} &= 1.625 \pm 0.013\ \text{(stat)}\ \pm 0.017\ \text{(syst) ps} \\
\tau_{B^0} &= 1.543 \pm 0.020\ \text{(stat)}\ \pm 0.033\ \text{(syst) ps} \\
\frac{\tau_{B^+}}{\tau_{B^0}} &= 1.051 \pm 0.019\ \text{(stat)}\ \pm 0.024\ \text{(syst)}
\end{aligned}$$

The same method was used to extract the average b-hadron lifetime from an unenhanced sample and resulted in:

$$\tau_b = 1.568 \pm 0.005\ \text{(stat)}\ \pm 0.009\ \text{(syst) ps}$$

These results confirm, with a significance of about 1.5 standard deviations, the predicted hierarchy between the B^+ and B^0 lifetimes. The result for the B^+ and B^0 lifetime ratio represents the most precise single measurement of this quantity from Z^0 resonance data (see Figure 8.3) and is also well compatible with a previous DELPHI measurement of this quantity from an inclusive topological vertex method [117]. The result of the average b-hadron mean lifetime is in good agreement with the most precise previous DELPHI publication on this subject [118]. It is currently the most precise extraction of τ_b worldwide and provides a significant improvement in the current world average (see Figure 8.4). Figures 8.1, 8.2, 8.3 and 8.4 show the experimental results from various collaborations together with the world average from the LEP lifetime working group. The results from this analysis are also shown in these plots.

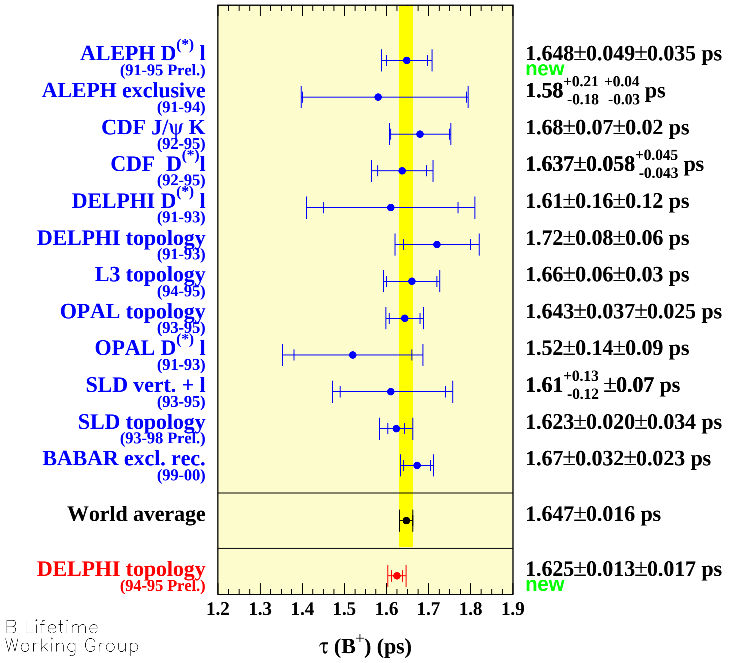

Figure 8.1: *Experimental results of the B^+ meson lifetime with error bars from various collaborations. Also the current world average value is shown and is compared with the result of this analysis.*

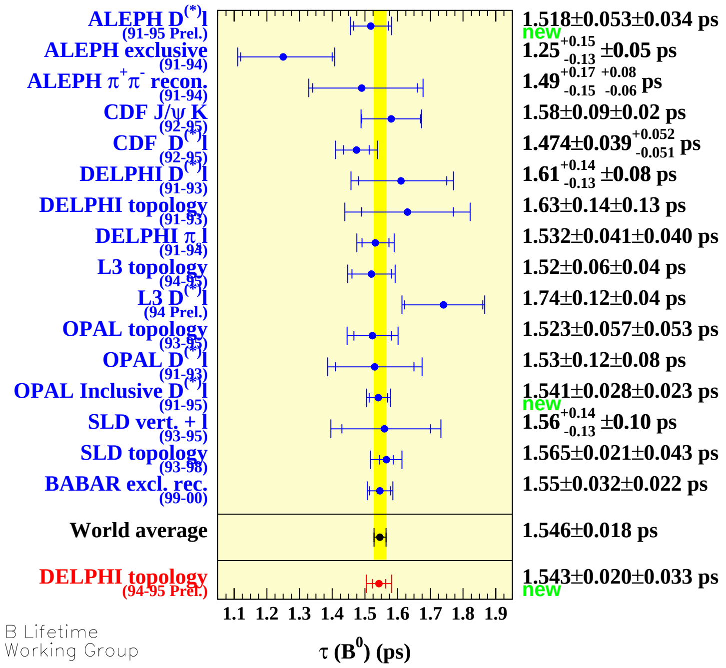

Figure 8.2: *Experimental results of the B^0 meson lifetime with error bars from various collaborations. Also the current world average value is shown and is compared with the result of this analysis.*

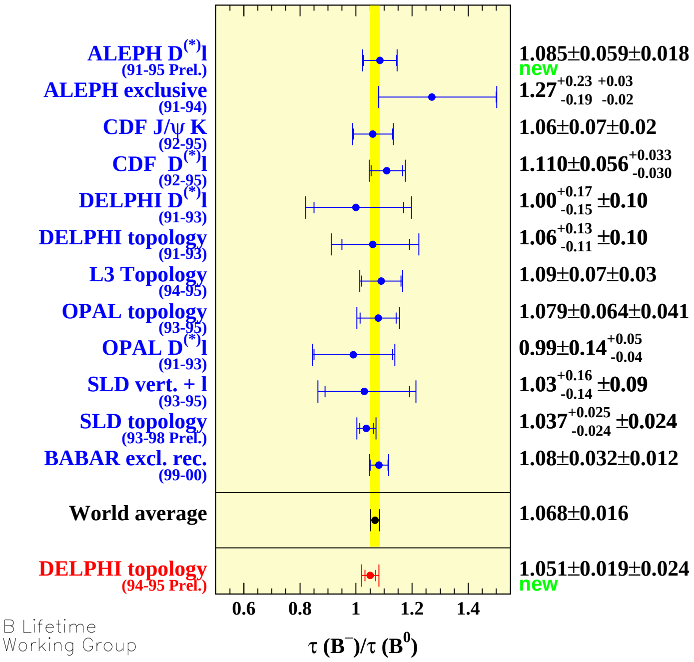

Figure 8.3: *Experimental results of the lifetime ratio of B^+ and B^0 mesons with error bars from various collaborations. Also the current world average value is shown and is compared with the result of this analysis.*

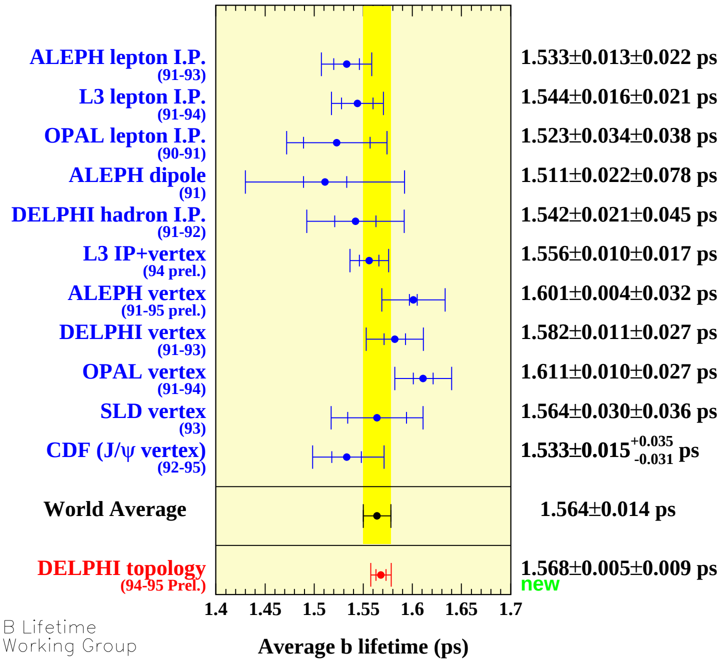

Figure 8.4: *Experimental results of the average b-hadron lifetime with error bars from various collaborations. Also the current world average value is shown and is compared with the result of this analysis.*

Bibliography

[1] *Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton-Nucleus Collisions*, Physical Review Letters 39, p. 252, (1977)

[2] L. M. Lederman, *The Upsilon Particle*, Scientific American, October 1978, vol. 239, no. 4, pp. 72-80

[3] W. Bartel et al., *UPPER LIMIT ON BEAUTY LIFETIME AND LOWER LIMIT ON WEAK MIXING ANGLES*, Phys.Lett.B114:71,1982

[4] G.S. Abrams et al., *MEASUREMENT OF THE LIFETIME OF BOTTOM HADRONS*, Phys.Rev.Lett.51:1316,1983

[5] W. Ash et al., *LIFETIME OF PARTICLES CONTAINING B QUARKS*, Phys.Rev.Lett.51:1022,1983

[6] The ALEPH Collaboration, *Measurement of the $\bar{B}^0$ and B^- Meson Lifetimes*, CERN-EP/2000-106

[7] G. Abbiendi et al., *Measurement of the B^0 Lifetime and Oscillation Frequency using $\bar{B}^0 \to D^{*+}l^-\bar{\nu}$ decays*, OPAL PR 317, CERN-EP/2000-90

[8] S. L. Glashow, Nucl. Phys. Rev. 96 (1954) 191

[9] S. Weinberg, Phys. Rev. Lett. 19 (1967) 1264

[10] A. Salam, Proc. 8th Nobel Symp. (1968) 367

[11] LEP Higgs Working Group, *Search for the Standard Model Higgs Boson at LEP*

[12] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. 10(1963), 531-532

[13] S. Glashow, J. Iliopoulos, L. Maiani, *Weak Interaction with Lepton-Hadron Symmetry.* Phys. Rev. D2, 1970

[14] J. J. Aubert et al., *Experimental Observation of a Heavy Particle J*, Phys. Rev. Lett. 33(1974). 1404-1406

[15] M. Kobayashi, K. Maskawa, *CP Violation in the renormalizable Theory of weak Interaction*, Prog. Theor. Phys. 49, 1973

[16] O. Nachtmann, *Phänomene und Konzepte der Elementarteilchenphysik*, Vieweg 1986

[17] T. Hebbeker, *Tests of Quantum Chromodynamics in Hadronic Decays of Z^0-Bosons produced in e^+e^- Annihilation*, Phys. Rep. 217, 1992

[18] Particle Data Group (PDG), *Review of Particle Physics*, Eur. Phys. J. D66, 010001 2002

[19] G. Altarelli und G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. B126, 1977, 298

[20] T. Sjöstrand, *High-Energy-Physics Event Generation with PYTHIA 5.7 and JETSET 7.4*, Computer Physics Communications 82, 1994, 74; T. Sjöstrand, *PYTHIA 5.7 and JETSET 7.4 Physics and Manual*, hep-ph/9508391

[21] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Physics Reports 97, 1983, 31

[22] G. Machesini und B. R. Webber, *Simulation of QCD Jets Including Soft Gluon Interference*, Nucl. Phys. B238, 1984, 1

[23] C. Peterson et al., *Scaling Violations in Inclusive e^+e^- Annihilation Spectra*, Phys. Rev. D27, 1983, 105

[24] R.D. Field und R.P. Feynman, *A Parametrization of the Properties of Quark Jets*,Nucl. Phys. B136, 1978, 1

[25] G. 't Hooft, *Renormalizable Lagrangians for massive Yang-Mills fields*, Nucl. Phys. B35, 1971, 167-188

[26] CDF Collaboration, *Observation of Top Quark Production in $p\bar{p}$ Collisions with the CDF Detector at Fermilab*, FERMILAB-PUB-95/022-E

[27] OPAL-Kollaboration, R. Akers et al., *Measurement of the Multiplicity of Charm Quark Pairs from Gluons in Hadronic Z^0 Decays*, Phys. Lett. B353, 1995, 595

[28] DELPHI-Kollaboration, P. Abreu et al., *Measurement of the Multiplicity of Gluons Splitting to Bottom Quark Pairs in Hadronic Z^0 Decays*, Phys. Lett. B405, 1997

[29] ALEPH-Kollaboration, *Measurement of the Gluon Splitting Rate to* $b\bar{b}$ *in Hadronic Z Decays*, Contribution to the International Europhysics Conference on High Energy Physics 1997, Jerusalem; Ref #606

[30] DELPHI-Kollaboration, *Study of Charm Meson Production in* Z^0 *Decays and Measurement of* $\Gamma_{charm}/\Gamma_{hads}$, Contribution to the International Europhysics Conference on High Energy Physics 1995, Bruxelles; Ref #0557 und DELPHI Note 95-101 PHYS 536

[31] DELPHI-Kollaboration, *Inclusive Measurement of the b Fragmentation Function*, Contribution to the International Europhysics Conference on High Energy Physics 1995, Bruxelles; Ref #0560 und DELPHI Note 95-103 PHYS 538

[32] N. Isgur and M. B. Wise, Phys. Lett. B232 (1989) 113;
M. B. Wise, Proc *XVI Int. Symp. on Lepton and Photon Interaction*, Ithaca, N. Y. 1993, Ed. P. Drell, D. Rubin, p. 253.

[33] N. Isgur and M. B. Wise, *'Heavy Quark Symmetry' in advanced series on directions in High Energy Physics* Vol. 10 'Heavy Flavours' World Scientific 1992, Ed. A. J. Buras, M. Lindner, p. 253.

[34] N. Isgur and M. B. Wise, *Heavy Quark Symmetry, B Decays*, World Scientific 1992 2nd Edition, Ed. S. Stone, p. 158.

[35] A. F. Falk et. al. Nucl Phys. B357 (1991) 185.

[36] *B Decays*, World Scientific 1992 2nd Edition, Ed. S. Stone

[37] G. Altarelli, N. Cabibbo, G. Corbo, L. Maiani and G. Martinelli, Nucl. Phys. B208 (1982) 365.

[38] A. Lenz, a talk given at the Workshop on the CKM Unitarity Triangle, CERN, Feb. 13th-16th 2002, http://ckm-workshop.web.cern.ch/ckm-workshop/

[39] ARGUS Collab., H. Albrecht et al., Phys. Lett. B249 (1990) 360;
E. H. Thorndike, Proc. 1985 *Int. Sym. on Lepton and Photon Interactions at High Energies*, edited by M. Konumu and K. Takahashim, Kyoto 1986, p. 406.

[40] C. S. Kim, A. D. Martin, Phys. Lett. B225 (1989) 186.

[41] A. Ali, E. Picturinen, Nucl. Phys. B154 (1979) 519.

[42] M. V. Danilov, *Heavy Flavour Physics, plenary session talk presented at the Lepton-Photon Symposium and Europhysics Conference on High Energy Physics*, Geneva, 1991.

[43] The LEP Main Ring Accelerating Structure, CERN 89-09.

[44] DELPHI-Kollaboration, P. Aarnio et al., *The DELPHI Detector at LEP*, Nucl. Instr. and Meth. A303(1991), 233

[45] DELPHI-Kollaboration, P. Abreu et al., *Performance of the DELPHI Detector*, Nucl. Instr. and Meth. A378(1996), 57

[46] DELPHI-Kollaboration, P. Abreu et al., *Performance of the DELPHI Detector*, DELPHI-Note 95-112, PHYS 547, Brussels EPS95, Ref. eps0764, 1995

[47] G. van Apeldoorn et al., *Performance of the Barrel Ring Imaging Cherenkov Counter of DELPHI*, DELPHI-Note 94-18 RICH 61 (1994)

[48] *DELPHI Data Analysis Program (DELANA) User's Guide*, DELPHI Note (1989), no. 44, PROG 137

[49] Y. Sacquin, *Description of the DELPHI DST Content*, DELPHI Note 161 (1994), no. PROG 210

[50] S. Quinton et al., *An Overview of the First and Second Level Trigger of DELPHI*, IEEE Trans. Nucl. Sci. 36 (1989), 390

[51] V. Bocci et al., *Basic concepts and architectural details of the DELPHI trigger system*, IEEE Trans. Nucl. Sci. 42 (1995), 837

[52] W. Adam et al., *Design and performance of the DELPHI data acquisition system*, IEEE Trans. Nucl. Sci. 39 (1992), 166

[53] DELPHI Collab., *The Design and Operation of the Slow Controls for the DELPHI-Experiment at LEP*, DELPHI-Note 94-14 DAS-151 (1994)

[54] *The ZEBRA System*, CERN Program Library entries Q100 and Q101, CERN, Genf, 1993

[55] *DELSIM, DELPHI Event Generation and Detector Simulation User's Guide*, DELPHI Note 89-67 PROG 142; *DELSIM, DELPHI Event Generation and Detector Simulation - Reference Manual*, DELPHI Note 89-68 PROG 143

[56] T. Sjöstrand, Computer Phys. Comm. 39 (1986) 347;
T. Sjöstrand, M. Bengtsson, Computer Phys. Comm. 43 (1987) 367;
T. Sjöstrand, *High-Energy-Physics Event Generation with PYTHIA 5.7 and JETSET 7.4*, Computer Phys. Comm. 82 (1994) 74;
T. Sjs̆trand, PYTHIA 5.7 and JETSET 7.4 Physics and Manual, **hep-ph/9508391**.

[57] G. Marchesini et al., *HERWIG: A Monte Carlo Event Generator for Simulating Hadron Emission Reactions with Interfering Gluons*, Computer Physics Communications 67(1992) 465

[58] L. Lonnblad, *ARIADNE Version 4: A Program for Simulation of QCD Cascades Implementing the Color Dipole Model*, Computer Physics Communications 71(1992) 15

[59] R. Fisher, *The use of multiple measurements in taxonomic problems*, Annals of Eugenics 7 (1936), 197-188

[60] G. Cowan, Statistical Data Analysis, oxford Science Publications, 1998

[61] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the theory of neural computation*, Studies in the sciences of complexity, Addison-Wesley Publishing

[62] B. A. Pearlmutter, *Dynamic Recurrent Neural Networks*, Tech. Rep. Nos. CMU-CS-90-196 (1990), Pittsburgh, PA 15213: School of Computer Science, Carnegie Mellon University

[63] J. A. Anderson, *Neural models with cognitive implications*, In D. LaBerge & S. J. Samuels (Eds.), Basic Processes in Reading Perception and Comprehension Models (p. 27-90), Hillsdale (1977), NJ: Erlbaum

[64] T. Kohonen, *Associative Memory: A System-Theoretical Approach*, Springer Verlag (1977)

[65] J. J. Hopfield, *Neurons with graded response have computational properties like those of two-state neurons*, Proceedings of the National Academy of Science (1982), 79, 2554-2558

[66] K. Hornik, M. Stinchcombe & H. White, *Multilayer feedforward networks are universal approximators*, Neural Networks (1989), 2(5), 359-366

[67] K.-I. Funahashi, *On the approximate realization of continuous mappings by neural networks*, Neural Networks (1989), 2(3), 193-192

[68] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, Mathematics of Control, Signals and Systems (1989), 2(4), 303-314

[69] E. J. Hartman, J. D. Keeler & J. M. Kowalski, *Layered neural networks with Gaussian hidden units as universal approximations*, Neural Computation (1990), 2(2), 210-215

[70] D. Anderson, *Supervised Learning of Probability Distributions by Neural Networks*, Neural Information Processing Systems, American Institute of Physics, N. Y., 1988

[71] J. J. Hopfield, *Learning Algorithms and Probability Distributions in Feed-Forward and Feed-back Networks*, Volume 84 of Proceedings of the National Academy of Science, USA, 1987

[72] S. Solla, E. Levin and M. Fleischer, *Accelerated learning in layered neural networks*, Complex Systems, 2:625-639, 1988

[73] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipies in C: Thev art of Scientific Programming*, Cambridge University Press, Cambridge, england, 1988

[74] E. Polak, *Computational Methods in Optimization*, New York: Academic Press, 1971

[75] C. M. Bishop, E. Bienenstock, R. Doursat, *Neural Networks and the Bias/Variance Dilemma*, Neural Computation (1992), 4, 1-58

[76] A. S. Weigend, D. E. Rumelhart, B. A. Huberman, *Generalization by Weight-Elimination with Application to Forecasting*, In: R. P. Lippmann, J. Moody & D. S. Touretzky (eds.), advances in Neural Information Processing Systems 3 (1991), San Mateo, CA: Morgan Kaufmann

[77] D. S. Broomhead and D. Lowe, *Multivariable Function Interpolation and Adaptive Networks*, Complex Systems, 2:321-355, 1988

[78] J. E. Moody and C. J. Darken, *Fast Learning in Networks of Locally-Tuned Processing Units*, Neural Computation, 1:281-294, 1989

[79] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995

[80] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, Inc., 2nd edition, 1987

[81] C. Cortez and V. Vapnik, *Support vector networks*, Machine Learning, 20:273-297, 1995.

[82] C. J. C. Burges, *A Tutorial on Support Vector Machines for Pattern Recognition*, Kluwer Academic PUblishers, Boston, 1-43

[83] M. Feindt, *NeuroBayes - A Neural Bayesian Estimator for Conditional Probability Densities*, Institut für Experimentelle Kernphysik, Internal Note, Karlsruhe, 2001

[84] C. M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995

[85] M. J. L. Orr, *Regularization in the Selection of Radial Basis Function Centers*, Neural Computation, 7(3):606-623, 1995

[86] F. James, J. Hoogland, R. Kleiss, *Multidimensional sampling for simulation and integrations: measures, discrepancies and quasi-random numbers*, NIKHEF 96-017, 1996

[87] A. R. Krommer, C. W. Überhuber, *Computational Integration*, SIAM 1998

[88] E. Hlawka, *Funktionen von beschränkter Variation in der Theorie der Gleichverteilung*, Ann. Math. Pura Appl. 54, 325-333 (1961)

[89] J. Bossert, *quasi-Zufallszahlen und ihre Anwendung in der Hochenergiephysik*, IEKP-KA 2001-11 (2000)

[90] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM (1992), Philadelphia

[91] J. H. Halton, *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals*, Num. Math. 2 (1960), 84-90

[92] T. Allmendinger, *Neutral B Meson Oscillations in Inclusive Samples with the DELPHI Detector at LEP*, IEKP-KA 2002-4 (2002)

[93] V. Blobel, E. Lohrmann, Statistische und numerische Methoden der Datenanalyse, B. G. Teubner Stuttgart, Leipzig 1998

[94] Z. Albrecht, T. Allmendinger, G. Barker, M. Feindt, C. Haag and M. Moch, *BSAURUS: A package for inclusive B reconstruction in DELPHI*, 2001,DELPHI-2000-069-PHYS-868 hep-ex/0102001

[95] G. Borisov and C. Mariotti, *Fine tuning of track impact parameter resolution of the DELPHI detector*, DELPHI Note (1995), No. 142, PHYS 567

[96] *Fine tuning of track impact parameter resolution of the DELPHI detector*, Nucl. Inst. Meth. A372 (1996), 181-187

[97] W. J. Murray, *Measurement of the Beam Position in DELPHI*, DELPHI Note 96-6 PHYS 590

[98] W. J. Murray, *Improved B tagging Using Impact Parameters*, DELPHI Note 95-167 PHYS 581

[99] G. V. Borisov, *Lifetime tag of events with B hadrons with the DELPHI detector*, IFVE-94-98

[100] G. V. Borisov, *Combined b-tagging*, DELPHI Note 97-94 PHYS 716

[101] T. Sjöstrand, *Computer Physics Communications*, 82 (1994) 74

[102] Z. Albrecht, M. Feindt, M. Moch, *MACRIB - High Efficiency, High Purity Hadron Identification for DELPHI*, DELPHI internal Note, 99-150 RICH 95 (1999)

[103] M. Feindt, C. Kreuter, O. Podobrin, *ELEPHANT reference manual*, DELPHI internal note, 96-82 PROG 217 (1996)

[104] H. Herr et al, *The Results of the combined beam test of the DELPHI hadron calorimeter, the forward electromagnetic calorimeter and barrel muon chambers (π^+, e^+ runs)*, HU-SEFT-1990-07

[105] J. Buytaert et al., *The Forward Muon Detector of the DELPHI experiment at LEP*, Nucl. Inst. Meth. A310 (1991), 596-606

[106] K. Münich et al., *Determination of the forward-backward asymmetry of b-quarks using inclusive charge reconstruction and lifetime tagging at LEP I*, DELPHI conference report, 2000-102 CONF 401

[107] D. E. Groom et al., Eur. Phys. J. C15 (2000) 1

[108] *The B-Oscillations Working Group*,
http://lepbosc.web.cern.ch/LEPBOSC/.RoscWG

[109] *B* Lifetime Group' *Averges of B Hadron Lifetimes*, July 2001,
http://lepbosc.web.cern.ch/LEPBOSC/lifetimes/lepblife_page2.html

[110] G. Barker, M. Feindt, U. Kerzel, L. Ramler, *A Study of the b-Quark Fragmentation Function with the DELPHI Detector at LEP I*, contribution to ICHEP 2002, Amsterdam, DELPHI 2002-069 CONF 603

[111] G. V. Borisov, C. Mariotti, Nucl. Inst. Meth. A372 (1996) 181

[112] C. Schwanda, *Inclusive b Decays to Wrong Sign Charmed Mesons*, DELPHI conference report, 2001-053 CONF 481

[113] MARK III Collaboration, D. Coffman et al., Phys. Lett. B263 (1991) 135

[114] T. Scheidle, *Bestimmung der b- und c-Quark Effizienz im DELPHI Detektor anhand einer 'DoubleTag' Methode*, IEKP-KA 2002-5 (2002)

[115] DELPHI Collaboration, P. Abreu et al. *Measurement of the Charged Particle Multiplicity of Weakly Decaying B Hadrons*, CERN-PPE-97-113, Aug 1997. 17pp, Phys.Lett.B416:233-246,1998

[116] A. H. Hoang et al., *B decay and the Υ mass*, UCSD/PTH 98-31, hep-ph/9809423

[117] DELPHI Collaboration, P. Abreu et al., Z. Phys. C68 (1995) 363

[118] DELPHI Collaboration, P. Abreu et al., Phys. Lett. B377 (1996) 195

Acknowledgements

This thesis would not have been possible without the help of many people. First I would like to thank my supervisor Michael Feindt, for his advice and continuous support and providing me with the possibility to work in the Karlsruhe DELPHI group. My thank goes also to Günther Quast for the co-supervision of the thesis and the careful reading.

Thanks go also to the Institute of Experimental Nuclear Physics in Karlsruhe and the Delphi Experiment for giving me the opportunity to undertake this research. I would like to thank also the *Graduiertenkolleg für Teilchen- und Astrophysik* for providing funding in the form of a graduate scholarship.

The many people in Karlsruhe who deserve my thanks for their help and friendship during my time there include: Dr. Gary Barker for writing the paper draft, for providing a running version of Bsaurus and for helping me to improve my English, Thorsten Scheidle for being a nice room mate and for discussions which were not only related to physics, Marcel Stanitzki and Jens Rehn for the time we shared together in Karlsruhe, Markus Moch for his careful reading of the german summary, Zoltan Albrecht for the joint preparation for the examination and Dr. Klaus Rabbertz for the careful reading of a previous draft.

Thanks go also to the Delphi B Mixing and lifetime group at CERN for their advice and interest.

Special thanks belongs to the Admin team Dr. Patrick Schemitz, Yves Kemp and Ulrich Kerzel for keeping the computer cluster running most of the time.

I would like to thank all members of the institute which have not been named explicitely for the nice atmosphere.

Most important of all, I thank my wife Susanne and her parents for their love and support in all the years. This work would not have been possible without their help. Finally I would also thank my children Laron and Finola for enriching my life. This thesis is dedicated to them.