Universität Karlsruhe (TH)

IEKP-KA/2003-16

Analyse von orbital angeregten B-Mesonen

Zoltán Albrecht

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

DISSERTATION

Tag der mündlichen Prüfung: 27.06.2003
Referent: Prof. Dr. M. Feindt, Institut für Experimentelle Kernphysik
Korreferent: Prof. Dr. G. Quast, Institut für Experimentelle Kernphysik

Contents

Kapitel 1

Deutsche Zusammenfassung

In diesem Kapitel wird die Analyse angeregter B-Mesonen zusammengefasst. Die Spektroskopie angeregter B-Mesonen beinhaltet die Messung der Masse der Zustände und die Bestimmung ihrer Erzeugungsrate. Die Motivation der Arbeit ist gegeben durch die enormen Entwicklungen auf dem Gebiet der statistischen Methoden und durch die verbesserte Spurrekonstruktion bei DELPHI. Die Analyse zielt auf die verbesserte Vermessung der B_u^{**}- und B_d^{**}-Mesonen, die bereits von DELPHI und anderen Kollaborationen analysiert wurden, sowie auf den Nachweis der Existenz von B_s^{**}-Mesonen, über die noch keine gesicherten Daten existieren. Die Anreicherung der angeregten Zustände mittels Neuronaler Netzwerke steht im Mittelpunkt der Arbeit.

1.1 Theoretischer Hintergrund

Der theoretische Hintergrund der Arbeit wird durch die Effektive Theorie Schwerer Quarks gegeben. Eine kurze Einführung soll einen Einblick in die vorhergesagten Eigenschaften der B^{**}-Mesonen geben.

1.1.1 Erzeugung von Hadronen bei LEP

Am LEP-Speicherring werden e^- und e^+ der Gesamtenergie 91,2 GeV zur Kollision gebracht. Dabei werden über den Austausch von einem Z^0-Boson Lepton- ($l\bar{l}$) oder Quark-Paare ($q\bar{q}$) erzeugt. Die für die Analyse interessanten Ereignisse beinhalten die Erzeugung von $b\bar{b}$-Paaren. In dieser sogenannten elektro-schwachen Phase werden in ca. 15% der Fälle $b\bar{b}$-Quarks produziert. In der darauf folgenden perturbativen Phase werden hochenergetische Gluonen und weitere Quarks abgestrahlt. Dieser Prozess wiederholt sich bis zur Bildung von Hadronen. Eine geschlossene theoretische Beschreibung der Prozesse bis hin zur Bildung der Hadronen ist jedoch nicht möglich, da α_s, die starke Kopplungskonstante, gegen 1 tendiert, die Theorie aber nur

für $\alpha_s \ll 1$ funktioniert. Diese Phase, genannt Fragmentation, kann nur mit Hilfe von Modellen beschrieben werden. Bei DELPHI ist es die String-Fragmentation, die Verwendung in der Simulation findet. Die produzierten Hadronen verlassen den Wechselwirkungsort in typischen Formationen, den Jets.

1.1.2 Effektive Theorie Schwerer Quarks

Die Wechselwirkung der Quarks wird über die Quantenchromodynamik (QCD) beschrieben. Freie Quarks sind nicht beobachtbar, sondern in Mesonen und Baryonen gebunden. Mesonen sind gebundene Zustände von einem Quark und einem Anti-Quark. Baryonen sind aus drei Quarks oder drei Anti-Quarks aufgebaut.

Der nicht-Abelsche Charakter der QCD stellt bei der theoretischen Vorhersage der Spektren von gebundenen Hadronen die QCD vor enorme Schwierigkeiten. Aufgrund der relativ großen Masse von b-Quarks kann eine Vereinfachung im Falle der B-Mesonen, d.h. einem gebundenen Zustand aus einem $\bar{b}$-Quark und einem beliebigen leichten Quark q, erzielt werden. Die Eigenschaften von B-Mesonen können mit den Eigenschaften eines Wasserstoff-Atoms verglichen werden: ein schwerer, unbeweglicher Kern in der Mitte, ein leichtes Quark in der Hülle und ein vom Ort des Kerns ausgehendes stationäres Potential. Die Theorie, die sich diese Eigenschaft zunutze macht, ist die Effektive Theorie Schwerer Quarks. Das leichte Quark spürt ein bindendes stationäres Potential, tritt aber ansonsten mit dem schweren $\bar{b}$-Quark nicht in Wechselwirkung. Die "chemischen" Eigenschaften des Mesons sind in dieser Näherung, der Näherung unendlicher $\bar{b}$-Masse, vom Spin und Flavour des $\bar{b}$-Quarks genauso wenig abhängig wie die chemischen Eigenschaften verschiedener Isotope von der tatsächlichen Anzahl der Neutronen im Kern.

1.1.3 Eigenschaften der B-Mesonen

Die Eigenschaften von B-Mesonen werden durch ihre Quantenzahlen, die die Auswahlregel bei Zerfällen bestimmen, sowie durch ihre Masse beschrieben. Folgt man der Idee der HQET, so kann man die Quantenzahlen von B-Mesonen leicht herleiten.

B-Mesonen im Grundzustand besitzen den Drehimpuls $L = 0$. In diesem Fall zeigt die z-Komponente des Spins des leichten Quarks in positive oder negative z-Richtung. Zusammen mit dem Spin des $\bar{b}$-Quarks ergibt sich, je nach Orientierung, ein Spin-1-Zustand, Vektor-Meson B^* genannt, mit $J^P = 1^-$, bzw. ein Spin-0-Zustand, pseudoskalares Meson B genannt, mit $J^P = 0^-$.

Orbital angeregte B-Mesonen mit Drehimpuls $L = 1$ werden B^{**}-Mesonen genannt. Der Spin des leichten Quarks ergibt mit diesem Drehim-

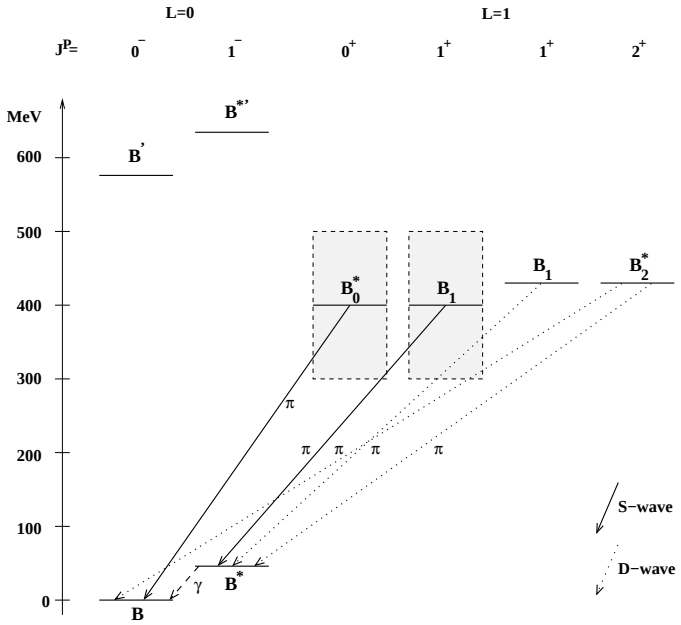

Abbildung 1.1: *Das Spektrum angeregter B-Mesonen. Die Skala ist bezogen auf die Masse des Grundzustandes B mit einer Masse von $m_B = 5.2789\ GeV/c^2$. Die Übergangslinien stellen hadronische Zerfälle in den Grundzustand dar. Diese erfolgen für $B^{**}_{u,d}$-Mesonen über Pionen und für B^{**}_s-Mesonen über Kaonen.*

puls $J_q = 1/2$ oder $J_q = 3/2$. Zusammen mit dem Spin des schweren Quarks ergeben sich also vier in Dubletts gruppierte Zustände mit den Quantenzahlen $(0^+,\ 1^+)$ und $(1^+,\ 2^+)$. Die angeregten Zustände zerstrahlen über das Aussenden eines Pions oder eines Kaons in die Grundzustände, wobei die erlaubten Übergänge durch die Quantenzahlen bestimmt sind. Der Zerfall erfolgt dabei für das Dublett mit $J_q = 1/2$ über eine S-Welle und für das Dublett mit $J_q = 3/2$ über eine D-Welle. Das Überlappintegral bestimmt die Breite des Zerfalls. So ist die Zerfallsbreite des ersten Dubletts groß und die Zerfallsbreite des zweiten klein.

Das Spektrum der orbital angeregten B-Mesonen ist in Abbildung 1.1 gezeigt. Die Vorhersage der Massen kann im Rahmen der HQET erfolgen, wenn die Massen der B-Mesonen mit den Massen der D-Mesonen in Relation gesetzt werden. Dynamische Aussagen können von der HQET nicht getroffen werden, da keine implizite Annahme über das Potential gemacht wird. Andere Modelle, wie z.B. Rechnungen auf dem Gitter, können auch dynamische Vorhersagen machen.

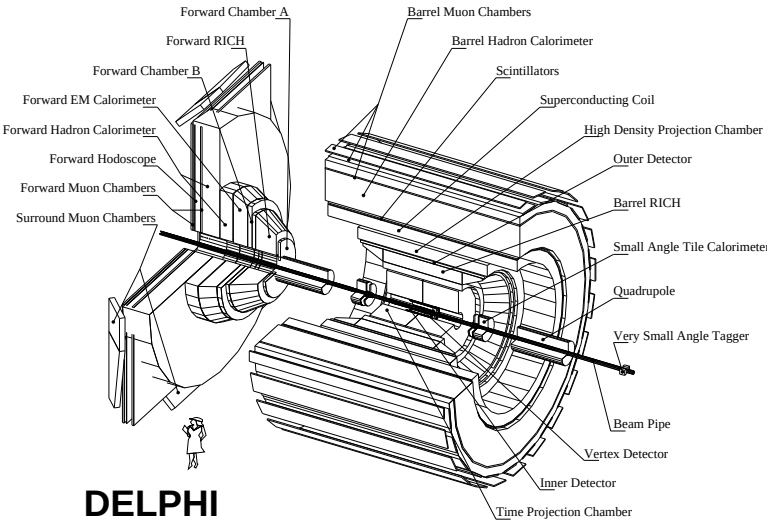

Abbildung 1.2: *Schematische Darstellung des DELPHI-Detektors.*

1.2 Das Experiment

Die analysierten Daten wurden mit dem DELPHI-Detektor aufgenommen, der am LEP-Speicherring (CERN) bei Genf betrieben wurde. Die zur Kollision gebrachten Elektronen und Positronen wurden in dem 27 km langen Speicherring rund 100 m unter der Erde zur Verfügung gestellt. LEP wurde von 1989 bis 2000 betrieben, zuerst bei der Z^0-Resonanzenergie von 91,2 GeV und später in der LEP II-Phase mit immer weiter gesteigerter Energie.

Der DELPHI-Detektor (Abbildung 1.2) war einer von vier Detektoren, die an den Wechselwirkungspunkten installiert waren. Den DELPHI-Detektor zeichnete die beinahe 4π-Winkelabdeckung sowie die hohe Granularität aus, die auch das Aufzeichnen von dichten Teilchenspuren ermöglicht. Der Aufbau des Detektors folgte dem Zwiebel-Prinzip, d.h. die Unterdetektoren waren in Schichten angeordnet. Dem Wechselwirkungsort am nächsten waren sehr präzise Spurenkammern, gefolgt von Detektoren für die Teilchenidentifikation und die Energiemessung. Ganz außen waren die Myonkammern angebracht. Für die Impulsmessung von geladenen Teilchen besaß der Detektor ein 1,2 Tesla starkes Magnetfeld, das von einer supraleitenden Spule erzeugt wurde.

Die für die hier vorgestellte Analyse wichtigsten Detektorkomponenten waren das Spurrekonstruktionssystem und der RICH zur Teilchenidentifikation. Zum Spurrekonstruktionssystem gehörte der Silizium-Vertex-Detektor (VD), der Inner-Detektor, die Zeit-Projektionskammer (TPC) und der Outer-Detektor. Messpunkte entlang der Trajektorien geladener Teil-

chen wurden off-line zu Spuren zusammengesetzt.

1.3 Analysewerkzeuge

1.3.1 Neuronale Netzwerke

Neuronale Netzwerke (NN) werden in der Analyse zu zweierlei Aufgaben eingesetzt: erstens zu Klassifikationszwecken und zweitens zur Vorhersage reellwertiger Größen. Der Einsatz von NN anstelle klassischer Diskriminierungstechniken ist bedingt durch die nicht-lineare Trennbarkeit der Signal- und Untergrundverteilungen. Basis der nichtlinearen Trennung ist die Verallgemeinerung der Stufenfunktion: der Sigmoid

$$g = \frac{1}{1 + e^{-a(\mathbf{x}-\mu)}} \ . \tag{1.1}$$

Über die Parameter a und μ können die Eigenschaften dieser Entscheidungsfunktion eingestellt werden. Im Limes $a \to \infty$ geht der Sigmoid in die bekannte Stufenfunktion über.

Neuronale Netzwerke beinhalten eine Verkettung von Neuronen, die über individuell eingestellte Gewichte verbunden sind. Ob das angesprochene Neuron feuert oder nicht, wird nicht binär entschieden (dies wäre der Fall mit einer Stufen-Entscheidungsfunktion), sondern der Grad der Anregung wird über den Funktionswert der Sigmoide bestimmt. Im Allgemeinen wird der Parameter a aller Sigmoide konstant gehalten. Die gewichtete Summe aller in ein Neuron hineinfließenden Werte bestimmt den Ort, an dem der Sigmoid ausgewertet wird. Den Prozess der Einstellung aller in einem Netzwerk vorhandenen Gewichte und Schwellen nennt man Netzwerk-Training.

1.3.2 Ereignis-Rekonstruktion

Die Analyse von B^{**}-Mesonen setzt die Rekonstruktion der zugehörigen Ereignisse voraus. Hierzu gehören:

- Multihadron-Selektion.
- Die Rekonstruktion des primären Vertizes (PV), d.h. des Wechselwirkungspunktes.
- Die Selektion von b-Ereignissen. Diese erfolgt im Wesentlichen über die Impaktparametermethode. B-Mesonen, die in b-Ereignissen entstehen, haben eine relativ hohe Lebensdauer, so dass ihre Zerfallsspuren nicht vom primären Vertex, sondern von einem davon um Millimeter versetzten Vertex stammen. Der Impaktparameter ist der kleinste Abstand, den eine Teilchenspur vom primären Vertex besitzt. Die Verteilung

der Impaktparameter der Spuren eines Ereignisses läßt b-Quarks erkennen, da sie in diesem Fall zu deutlich positiven Werten verschoben ist. Nicht-b-Ereignisse haben eine Impaktparameterverteilung um Null. Die Größe, die angibt, wie wahrscheinlich es ist, dass das betrachtete Ereignis ein b-Ereignis ist, nennt man b-tag.

- Die Zuordnung der Spuren einer Hemisphäre zum primären oder zum sekundären Vertex. Diese Entscheidung wird mittels eines Neuronalen Netzwerkes getroffen und basiert hauptsächlich auf der Rapidität einer Spur. Die Rapidität ist eine Lorentz-invariante Größe, die ein Maß für den Winkel zwischen den Spuren ist. Teilchen, die vom B-Zerfall stammen, haben eine hohe Rapidität, da sie im Schwerpunktsystem des B-Mesons isotrop zerfallen und deshalb nach dem Boost in das Laborsystem mit dessen Impuls fast parallel sind. Die im Vergleich mit dem B-Meson niederenergetischen Teilchen aus der Fragmentation sind mit dem Zerfall des B-Mesons nicht korreliert und haben deshalb eine niedrige Rapidität.

- Die Rekonstruktion des sekundären Vertizes (SV), d.h. des Ortes, an dem das b-Hadron zerfallen war, mit Hilfe des Vertexzuordnungsnetzes.

- Die Rekonstruktion der B-Energie und Richtung. Auch diese Größen beruhen auf der Ausgabe des Vertexzuordnungsnetzes, denn die Impulse der Teilchen einer Hemisphäre, deren Rapidität einen bestimmten Wert ($y > 1,6$) überschreitet, werden gemäß ihrer SV-Zugehörigkeit gewichtet aufsummiert.

- b-Hadron-Identifikation. Diese Aufgabe wird von einer Reihe von Neuronalen Netzwerken übernommen, und ihre Ausgabe bestimmt einerseits die Sorte des b-Hadrons, d.h. B^+, B^0, B_s oder Λ_b, und andererseits die Ladung des b-Quarks im Hadron, d.h. b mit Ladung $-1/3$ oder $\bar{b}$ mit Ladung $+1/3$.

Die erwähnten Punkte der Ereignis-Rekonstruktion sind im Programmpaket BSAURUS implementiert, das eine gemeinsame Entwicklung der DELPHI-Gruppe des Instituts ist.

1.3.3 Teilchenidentifikation

Der Teilchenidentifikation kommt in der Arbeit eine besondere Rolle zu, da die Identifikation von Kaonen essentiell für die Analyse von B_s^{**}-Mesonen ist. Da die Standard-DELPHI Kaonidentifikation nicht ausreichend war, wurde eine neue Methode entwickelt, die auf die Technik der Neuronalen Netzwerke beruht. Ebenso zur Identifikation von Protonen bei allen vorkommenden Impulsen werden Neuronale Netzwerke zur Verfügung gestellt.

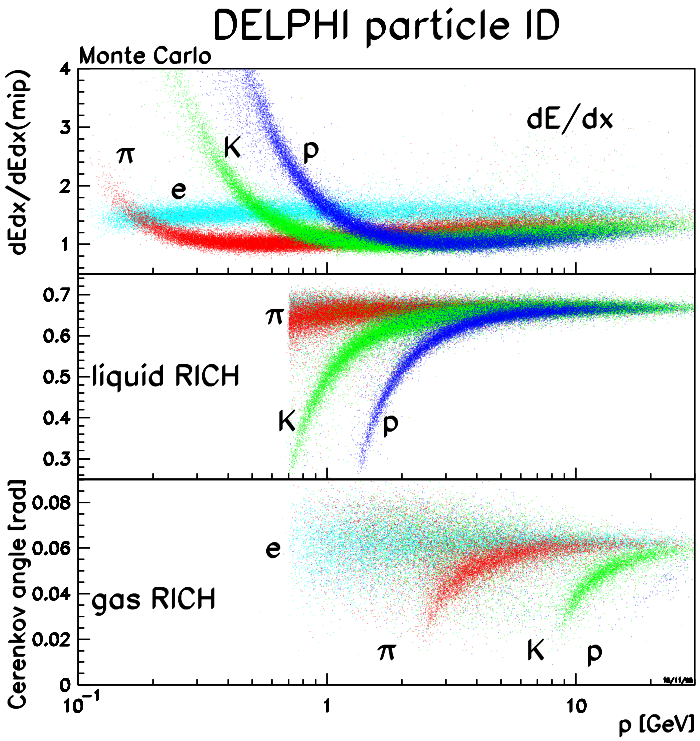

Abbildung 1.3: *Teilchenidentifikation in DELPHI.*

Teilchenidentifikation wird bei DELPHI mittels den RICH sowie der TPC und der VD Detektoren betrieben. Das Prinzip des RICH (Ring Imaging Cherenkov Counter) beruht hierbei auf dem Aussenden elektromagnetischer Strahlung beim Passieren geladener Teilchen durch Materie, sobald diese Teilchen die Lichtgeschwindigkeit in diesen Medien überschreiten. Die TPC und der VD liefern den spezifischen Energieverlust, den ein geladenes Teilchen in ihnen erlitten hat. Dieser Energieverlust beruht dabei auf dem Effekt der Ionisation und Anregung der Atome in dem Detektor und kann durch die Bethe-Bloch-Formel beschrieben werden. Die Messungen, die in den RICH-Detektoren und in der TPC gemacht wurden, sind in Abbildung 1.3 dargestellt. Nicht gezeigt sind die dE/dx-Messungen des Vertex-Detektors.

Eine optimale Kombination dieser Informationen mit Hilfe eines Neuronalen Netzwerkes führt zur Identifikation der Teilchensorte: Kaon oder Nicht-Kaon. Die Güte der Trennung ist in Abbildung 1.4 zu sehen. Links die Anreicherung von Kaonen bei großen positiven Werten sowie die Anreicherung von nicht-Kaonen bei großen negativen Werten. Rechts ist die Reinheit über die Effizienz aufgetragen. Eine perfekte Trennung könnte 100% Reinheit bei 100% Effizienz erzielen, würde also in der Ecke oben rechts des zweidimensionalen Histograms sitzen. Optimierte Methoden tendieren in diese

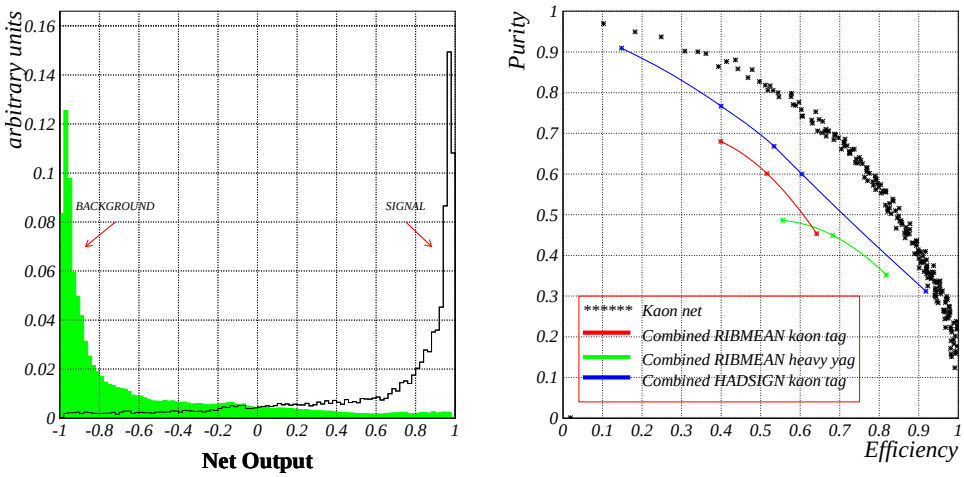

Abbildung 1.4: *DELPHI Kaonidentifikation.*

Richtung, und ihre Güte kann so durch den Vergleich mit klassischen Methoden abgeschätzt werden. Aus diesem Grund werden auch die Reinheiten und Effizienzen der standard DELPHI RIBMEAN und HADSIGN Algorithmen gezeigt, die eine deutlich geringere Leistungsfähigkeit besitzen.

1.4 Die experimentelle Methode

Der Zerfall eines angeregten Zustandes mit Masse M in ein B-Meson und ein leichtes π kann über die Relation $M = m_B + m_\pi + Q$ beschrieben werden. Dabei bezeichnet Q die bei der Reaktion freigewordene Energie. Stellt man diese Gleichung um, kommt man auf die zentrale Größe der Analyse:

$$Q(B^{**} \to B\pi) = M(B\pi) - m(B) - m(\pi) \tag{1.2}$$

Die Rekonstruktion des Q-Wertes wird für jede Kombination von Spuren mit dem in der Hemisphäre vorhandenen B-Kandidaten durchgeführt. Dabei treten folgende Schwierigkeiten auf:

- Die Zerfallsprodukte (B und π) sind durch den Boost in das Laborsystem fast parallel. Dadurch wirken sich schon relativ kleine Unsicherheiten in der Winkelmessung negativ auf die Auflösung aus.

- Das B-Meson wird nicht direkt, sondern über seine Zerfallsprodukte nachgewiesen. Durch individuelle Messfehler der einzelnen Spuren und die Unsicherheit der Zuordnung zum B-Zerfall werden große Fehler in die Richtungsmessung des B-Impulses eingeführt.

- Der primäre Zerfall erfolgt am e^+e^--Wechselwirkungspunkt, woher auch die Teilchen aus der Hadronisierung stammen. Diese Teilchen bilden einen Jet und können das Zerfallspion maskieren sowie die vorhin erwähnte Zuordnung von Spuren zum sekundären Vertex erschweren.

Der Rekonstruktion des Q-Wertes mit bestmöglicher Auflösung kommt dabei eine genauso wichtige Rolle zu wie der Anreicherung von Signal-Ereignissen in einer vorhandenen Verteilung. Eine Darstellung aller möglichen Kombinationen von B und π würde wegen der um Größenordnungen höheren Untergrundrate die Analyse von angeregten B-Mesonen unmöglich machen. Die Anreicherung wird dabei auf die kinematischen und ladungsbehafteten Korrelationen zwischen dem B und dem π sowie auf die exzellente Teilchenidentifikation von DELPHI gestützt. Der letzte Punkt ist besonders für B_s^{**}-Mesonen wichtig, deren Zerfall in den Kanal B^+K^- beobachtet wird. Da die Rate von Pionen ca. sieben mal höher ist als die der Kaonen, ist eine gute Kaonidentifikation absolut notwendig.

1.5 Die Analyse

Die Analyse beinhaltet die Rekonstruktion des Q-Wertes und die Anreicherung der Zustände. Die erhaltene Q-Wert-Verteilung muss mit den Erwartungen der Simulation verglichen werden. In der Simulation versucht man, die Ereignisse, die zum Untergrund in der Q-Wert-Verteilung beitragen, möglichst genau zu beschreiben, so dass übereinander aufgetragen der B^{**}-Anteil sichtbar wird. Dies ist einerseits um so leichter möglich, je besser die Auflösung ist und andererseits je kleiner der Untergrundanteil ist.

1.5.1 Q-Wert-Rekonstruktion

Spuren vom primären Vertex werden mit den rekonstruierten B-Mesonen kombiniert, und der zugehörige Q-Wert wird ermittelt. Dieser erste Schätzwert ist in zweierlei Hinsicht problematisch: erstens ist der Wert durch die Hinzunahme des B^{**}-Zerfallsteilchens zum Impuls des B-Mesons oft verfälscht. Zweitens hat man keine direkte Information über die Güte des Schätzwerts. Eine klassische Fehlerrechnung ist im Falle der Q-Wert-Berechnung wegen nicht-Gaußscher Fehler unmöglich.

Beide Probleme werden durch eine Korrektur behoben, die auf der Vorhersage eines Neuronalen Netzwerkes (QCorr) basiert. Das QCorr-Netz gibt die gesamte Wahrscheinlichkeitsdichte der Korrektur an. Ein solches Netzwerk wurde für jedes der analysierten B-Mesonen trainiert. So ergibt sich ein Q-Wert mit höchster Auflösung zusammen mit dem dazugehörigen Fehler, und dies ohne jegliche Verfälschung.

1.5.2 B^{**}-Anreicherung

Die Anreicherung von B^{**}-Ereignissen beruht auf einer Reihe von Neuronalen Netzwerken, deren Eingabevariablen sind:

- Variablen aus Hemisphären, die B^{**} enthalten. Diese Variablen beinhalten sowohl die Eigenschaften des schwach zerfallenden B-Mesons als auch Qualitätsinformationen über die Hemispäre im Allgemeinen. Diese sind im einzelnen:
  - Netzwerkinformation über die b-Hadron-Sorte.
  - Die 3-D Zerfallslänge zusammen mit ihrem Fehler.
  - Der b-Tag der Hemisphäre und des Ereignisses.
  - Die Hemisphärenqualität.
  - Der KaonNet-Wert der drei energiereichsten Fragmentationsteilchen. Diese Größe ist ein sehr empfindliches Maß für das Vorhandensein von B_s^{**}-Mesonen.
  - Die Anzahl von Fragmentations- und B-Zerfallsteilchen.
  - Die Anzahl von Kaonen in der Hemisphäre.
- Die Eigenschaften des Zerfalls-Pions/Kaons. Diese sind im einzelnen:
  - Die Ausgabe aller Teilchenidentifikationsnetzwerke.
  - Der Impuls und die Rapidität.
  - Der Winkel im B^{**}-Schwerpunktsystem. Da B^{**}-Mesonen nicht polarisiert sind, sind ihre Zerfallsprodukte in ihrem Schwerpunktsystem isotrop verteilt. Teilchen aus der Fragmentation hingegen scheinen entgegen der B^{**}-Flugrichtung gerichtet zu sein.
  - Die SV-Zugehörigkeit.
  - Die Spurqualität.
- Die kinematischen und ladungsbehafteten Korrelationen zwischen dem schwach zerfallenden B und dem Zerfallsteilchen:
  - Der Winkel α, den das primäre Teilchen und das B einschließen.
  - Die Ladungskorrelation zum Flavour des B-Mesons zum Produktions- und Zerfallszeitpunkt. Diese Information ist für neutrale B-Mesonen wichtig, die zwischen ihrer Produktion und ihrem Zerfall oszillieren können.
  - Die Korrelationen zur B_s- und B^0-Hemisphärenlikelihood. Diese sind die neutralen B-Mesonen, die oszillieren können.
  - Die Ladungskorrelation zum Jet.
  - Die Ladungskorrelation zu den in der Hemisphäre vorhandenen Pionen.

1.5.3 Ergebnisse

Untersucht wurden die angeregten Mesonen in den folgenden Zerfällen:

- $B_u^{**} \to B^0\pi^+$
- $B_d^{**} \to B^+\pi^-$
- $B_s^{**} \to B^+K^-$

Für diese Mesonen wurden individuelle Netzwerke zur Q-Wert-Rekonstruktion und zur Anreicherung trainiert. Im Falle der $B_{u,d}^{**}$-Mesonen konnte zur Q-Wert-Rekonstruktion dasselbe Netzwerk benutzt werden, da die Kinematik dieselbe ist und alle beteiligten Massen gleich sind.

Analyse der $B_{u,d}^{**}$-Mesonen

Da die Kinematik beider Zustände gleich ist, wurde im ersten Schritt ein generelles $B_{u,d}^{**}$-Netzwerk verwendet, dem nur Teilchen- und Hemisphärengrößen ohne Ladungsinformationen geliefert wurden. Nach der Anreicherung beider Zustände müssen B_u^{**}- und B_d^{**}-Mesonen voneinander getrennt werden. Dies erfolgt, indem die Ladungskorrelation zwischen den schwach zerfallenden B-Mesonen und den primären Teilchen ausgenutzt wird. Eine zusätzliche Anreicherung der Verteilungen in B^+- und B^0-Mesonen reduziert den Untergrund noch weiter. Abbildung 1.5 zeigt die Q-Wert-Verteilung der $B_{u,d}^{**}$-Mesonen aufgeteilt in vier Zonen. Horizontal wurden B^+-Mesonen von B^0-Mesonen getrennt und vertikal Teilchen mit negativer von Teilchen mit positiver Korrelation zum B-Meson. Erwartet wird für B_d^{**}-Mesonen ein Beitrag nur in der Verteilung links oben und für B_u^{**}-Mesonen in der Verteilung rechts unten. Dass auch noch Signalereignisse in den beiden anderen Verteilungen zu sehen sind, liegt an der endlichen Trennbarkeit des Problems. Insgesamt wurde für die B_d^{**} eine Reinheit von 30.7% und für die B_u^{**} eine Reinheit von 27.1% erreicht.

Die Verteilungen wurden simultan mit einer Gauß-Funktion für die schmalen B^{**}-Komponenten und einer Breit-Wigner-Funktion für eine zusätzliche breite Komponente gefittet. Die breite Komponente kann auch als ein radial angeregtes B-Meson mit $L = 0$ interpretiert werden. Für den Untergrund wurde die Vorhersage der Simulation mit einer linearen Korrektur für die Normierung und die Steigung verwendet. Die schmale und breite Signalkomponente ergeben eine $B_{u,d}^{**}$-Rate von

$$\begin{aligned} \frac{\sigma(B_{u,d}^{**})_{schmal}\mathrm{Br}(B^{**} \to B^{(*)}\pi)}{\sigma_b} &= 0.193 \pm 0.02(\mathrm{stat}) \pm 0.01(\mathrm{syst}) \\ \frac{\sigma(B_{u,d}^{**})_{breit}\mathrm{Br}(B^{**} \to B^{(*)}\pi)}{\sigma_b} &= 0.16 \pm 0.04(\mathrm{stat}) \pm 0.04(\mathrm{syst}) \end{aligned} \quad (1.3)$$

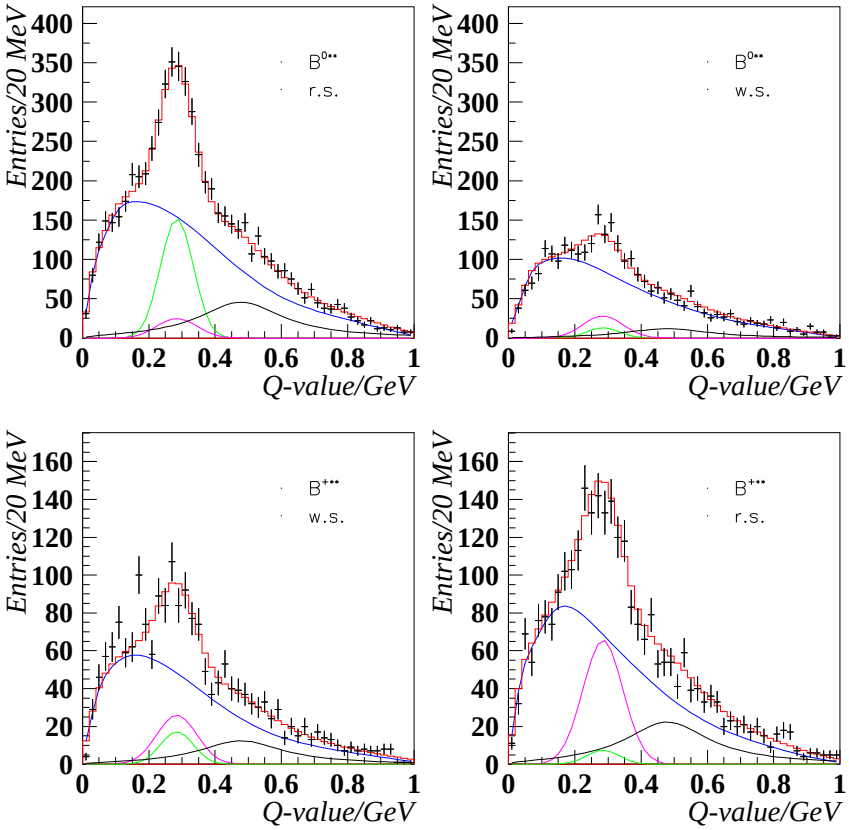

Abbildung 1.5: *Q-Wert-Verteilung der $B^{(*)}\pi$-Kandidaten.*

d.h. in circa 35% aller b-Ereignisse werden angeregte $B_{u,d}$-Mesonen produziert. Die Q-Werte der Resonanzen und ihre Breiten wurden auf die Werte

$$\begin{aligned} Q(B^{**}_{u,d} \to B^{(*)}\pi) &= 285 \pm 3(\text{stat}) \pm 7(\text{syst})\text{MeV/c}^2 \\ \sigma_Q(B^{**}_d) &= 49 \pm 5(\text{stat}) \pm 6(\text{syst})\text{MeV/c}^2 \\ \sigma_Q(B^{**}_u) &= 60 \pm 5(\text{stat}) \pm 6(\text{syst})\text{MeV/c}^2 \\ R &= 0.8 \pm 0.2(\text{stat}) \pm 0.2(\text{syst}) \\ \text{FWHM} &= 574 \pm 90(\text{stat}) \pm 80(\text{syst})\text{MeV/c}^2 \\ \text{MPV} &= 480 \pm 20(\text{stat}) \pm 18(\text{syst})\text{MeV/c}^2 \end{aligned} \tag{1.4}$$

bestimmt. R gibt das gefittete Breit-zu-Schmal-Verhältnis an, das in der Theorie zwischen 1:1, für den Fall daß Zustände gezählt werden, und 1:2, für den Fall daß Spins gezählt werden, schwanken kann.

Die Messung ergibt eine konstante und signifikante Rate für beide Fit-Komponenten. Daraus kann man ableiten, dass die schmalen Zustände der $B^{**}_{u,d}$-Mesonen und eine weitere $B^{**}_{u,d}$-Komponente bei großen Massen existrieren. Die schmalen Zustände sind durch die unzureichende Massenauflösung des Detektors nicht einzeln nachweisbar. Die Breite der an die Daten gefitteten Gauß-Funktion schließt aber das Vorhandensein aller von der HQET vorhergesagten schmalen Resonanzen innerhalb des Peaks nicht aus. Der breite Peak bei 500 MeV Q-Wert kann im Moment nicht eindeutig zugeordnet werden. Sowohl eine Interpretation als breiter $B^{**}_{u,d}$-Zustand als auch eine Interpretation als radiale Anregung können nicht ausgeschlossen werden. Dabei favorisieren die theoretischen Vorhersagen eine so hohe Masse der breiten Zustände nicht. Ausgeschlossen kann diese Situation allerdings nicht werden, zumal neue Erkenntnisse aus dem D-Sektor, gewonnen durch die BaBar- und CLEO-Kollaborationen, zeigen, dass die Spektroskopie der schweren Hadronen noch einige Überraschnungen parat hat. So wurde unter der DK-Schwelle eine $D_s(2320)$-Resonanz gefunden, deren Spin-Parität mit $J^P = 0^+$ verträglich ist und aus diesem Grund als breites D^{**}_s interpretiert werden kann. Erst weitere Messungen werden zur Klärung dieser Sachverhalte beitragen können. Dabei kann mit den DELPHI-Daten eine weitere Verbesserung der Ergebnisse erzielt werden, wenn eine dediziertere Behandlung der Signal- und Untergrund-Komponenten vorgenommen wird. In der Zukunft wird die LHCb-Kollaboration nach erfolgreicher Umsetzung ihres Detektor-Designs die Spektroskopie der D^{**} und B^{**} weiterführen können, unter Einbeziehung der neutralen Kanäle, die, wie bei BaBar und CLEO, wichtig für die Vollständigeit der Analyse sind.

Analyse der B^{**}_s-Mesonen

Nach der speziell auf B^{**}_s-Mesonen abgestimmten Q-Wert-Rekonstruktion wurden die Signalereignisse mit der in Abschnitt 1.5.2 beschriebenen Methode angereichert. Dabei wurden die diskriminierenden Variablen in ein einziges Netzwerk gegeben. Dies ist einerseits bei B^{**}_s-Mesonen möglich und andererseits erhöht es die Leistungsfähigkeit des Netzwerks, da so auch die Korrelationen richtig erfasst werden können.

Die Methode ergibt eine Signalanhäufung in der Q-Wert-Verteilung, die in Abbildung 1.6 zu sehen ist. Die Signalkomponente wurde mit der erwarteten Signalverteilung gefittet. Diese wurde in der Simulation ermittelt, um der Detektorauflösung bei dem betrachteten Anreicherungsgrad Rechnung zu tragen. Der Untergrund wurde analog zur B^{**}-Analyse behandelt. Die Signal-Rate wurde zu

$$\frac{\sigma(B^{**}_s)\ \mathrm{Br}(B^{**}_s \to BK)}{\sigma_b} = 0.0085 \pm 0.0023(\mathrm{stat}) \pm 0.0013(\mathrm{syst}) \tag{1.5}$$

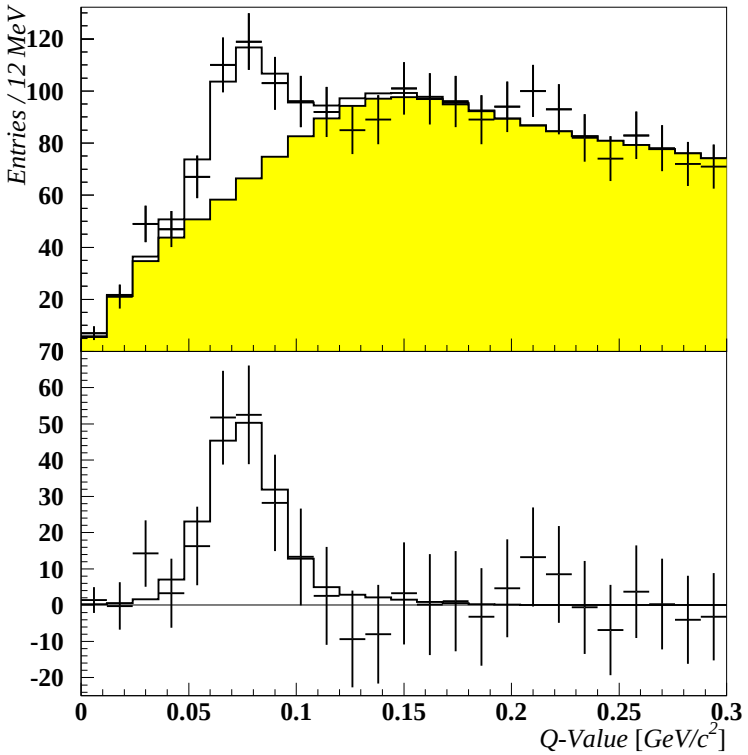

Abbildung 1.6: *B_s^{**} Q-Wert-Verteilung.*

bestimmt mit dem Q-Wert

$$Q(B_s^{**} \to BK) \quad = \quad 80.2 \pm 4.7(\text{stat}) \pm 4.2(\text{syst})\text{MeV/c}^2 \ . \tag{1.6}$$

Dies entspricht einer Masse von

$$Q(B_s^{**}) = 5853.2 \pm 4.7(\text{stat}) \pm 4.2(\text{syst})\text{MeV/c}^2 \ . \tag{1.7}$$

Die Interpretation dieses Ergebnisses hängt entscheidend davon ab wie groß die Signifikanz der Signal-Anhäufung ist. Dazu wurden zwei Tests durchgeführt: Erstens wurde versucht, die gesamte Verteilung in den Daten mit dem Untergrund-Spektrum alleine zu fitten. Das Ergebnis dieses Tests ergibt, dass man mit einer Signifikanz von 5.8σ (Standardabweichungen) ausschließen kann, dass die Daten durch den Untergrund beschrieben werden können. Zweitens wurde ein Test gemacht, ob die Signal-Anhäufung eine statistische Fluktuation sein kann. Dazu wurde die Signal-Komponente des vollen Fits entfernt und die Likelihood dafür berechnet, dass die Untergrund-Verteilung, ohne einen neuen Fit durchzuführen, die Daten beschreibt. Als

Ergebnis kann man feststellen, dass mit einer Signifikanz von 8.7σ ausgeschlossen werden kann, dass der Peak eine statistische Fluktuation ist.

Diesen Hinweis gibt auch die Konstanz der Ratenmessungen, die durchgeführt wurden, um den systematischen Fehler abzuschätzen. Hierfür wurden die Anreicherungs-Netzwerkschnitte variiert, zusammen mit Schnitten auf die Auflösung der Spuren sowie auf den Grad der Kaonanreicherung. Die Streuung der Werte um den Mittelwert ist hier als systematischer Fehler angegeben.

Der Vorhersage von HQET zufolge sollten vier angeregte Zustände existieren: zwei breite und zwei schmale. Die in den Daten gemessene Resonanz ist verträglich mit einer sehr schmalen Resonanz deren intrinsische Breite durch die endliche Detektorauflösung verdeckt wird. Die breiten B_s^{**}-Zustände können durch die hier vorgestellte Analysemethode nicht nachgewiesen werden. Die ist einerseits bedingt durch die unzureichende Kenntnis des Untergrundes und andererseits durch die Normierungs- und Steigunskorrektur des Monte Carlo-Untergrundes. Die schmalen Zustände sollten nachweisbar sein, sofern ihre Masse über der BK-Schwelle liegt. Da es keine Evidenz für eine weitere schmale Resonanz in den Daten gibt, werden obere Schranken für die nicht-beobachtete Zuständen gegeben. Der gemessene Peak kann auf zweierlei Art und Weise interpretiert werden:

- Erstens kann man den Peak den 1^+ und 2^+-Zuständen zuordnen, die beide in B^*K zerfallen. Da der 2^+-Zustand auch einen Zerfallskanal in BK besitzt, müsste dieser 46 MeV höher im Q-Wert sichtbar sein. Um eine obere Schranke angeben zu können, wurden die Vorhersagen der HQET sowie der vorhandene Phasenraum und die Detekor-Akzeptanz berücksichtigt. Die gemessene Rate ist bei 126 MeV Q-Wert kleiner als -0.0064% bei 95% Konfidenz. Die erwartete Rate liegt bei 0.51%. Dadurch kann man diese Möglichkeit ausschließen.

- Zweitens kann man den Peak dem 2^+-Zustand zuordnen, der in BK zerfällt. Analog zum ersten Fall müsste diesmal 46 MeV tiefer in Q-Wert eine Erhöhung sichtbar sein. Hier wurde für die Rate bei 95% Konfidenz eine obere Schranke von 0.142% bestimmt. Die erwartete Rate liegt bei 0.375%. Dadurch kann diese Möglichkeit in Betracht gezogen werden.

Das Nichtvorhandensein eines weiteren Peaks kann so interpretiert werden, dass der Zerfallskanal in $B^{(*)}K$ nicht der einzige ist und dass die Kopplung des 1^+-Zustandes an diese alternativen Kanäle stärker ist als dies beim 2^+-Zustand der Fall ist. Dies könnte auf die unterschielichen Massen zurückzuführen sein, die wegen der großen Nähe zur Schwelle entscheidend sein können. Eine weitere Möglichkeit ist, dass die besondere Form der Q-Wert-Verteilung, die wiederum eine Folge der nahen Schwelle ist, eine direkte Beobachtung der unter Umständen vorhandenen B_{s1}-Zustände verhindert.

Wie bei den $B^{**}_{u,d}$-Mesonen bereits erwähnt, kann man keine endgültigen Aussagen über die breiten Zustände machen, solange nicht auch die neutralen Zerfallskanäle untersucht wurden. Dies gilt insbesondere für den Zerfall $B^{**}_s \to B_s\pi^0$, der isospinverletzend ist und deshalb nur dann auftritt, wenn die B^{**}_s-Masse unter der BK-Schwelle sitzt.

Chapter 2

Introduction

Although heavy quark spectroscopy is now a rather mature subject, a number of interesting issues remain. In particular, the detailed properties of the excitation spectrum of heavy-light mesons (D, D_s, B, B_s) and their light hadronic transitions are yet to be fully understood. Central to theoretical predictions in the spectroscopy of b-hadron states is the utilization of an approximate heavy flavour symmetry arising from the fact that the mass of the b-quark is large relative to the scale Λ_{QCD}. In the resulting theoretical framework of Heavy Quark Effective Theory (HQET), the spin of the b-quark $\vec{s}_b$ in a b-hadron is conserved in production and decay processes *independently* of the total angular momentum of the light quark spectator system. Excitation energy levels are thus doublets in the *total spin* of the light-quark system j_q, with $\vec{j}_q = \vec{s}_q + \vec{L}$, and the different states can be labelled in terms of the spin-parity of the hadron J^P, with $\vec{J} = \vec{j}_b + \vec{j}_q$. Their relative production rate follows from spin state counting $(2J+1)$. For the dominant two-body decays, the $j_q = 1/2$ states can decay via an S-wave transition and their decay widths are expected to be broad in comparison to those of the $j_q = 3/2$ states which must decay via a D-wave transition. Unfortunately, HQET is not sufficient to determine the detailed properties of these states. Lattice gauge theory is the only existing technique which allows the systematic study of all the aspects of QCD in heavy-light systems. Models used to estimate the hadronic transitions from excited to ground states are more difficult because the light quark is subject to the full non-perturbative QCD dynamics.

The primary motivation for studying excited b-states is to test the predictions based on HQET and so make progress in understanding the true nature of the QCD potential. A further motivation comes from the observation [GNR93] that the large production rate of excited states can be useful in tagging the b-quark flavour (i.e. b or $\bar{b}$) of a b-hadron at the time of its production. This is already providing a powerful tool for flavour tagging [B$^+$98] which is an essential part of B^0 oscillation and CP-violation studies.

Experimentally, much of the excitation spectra remains to be observed. In the charm sector, the ground states and all six narrow states of the $j_q = 3/2$ doublet for each quark content $(c\bar{u}, c\bar{d}, c\bar{s})$ are well established (e.g. [B$^+$94]), whereas the wide $L = 1$ states are hard to measure and have not been clearly identified. In the beauty sector, the ground states are also well established with exception of the B_s^* meson, which has been observed at the $\Upsilon(5s)$ resonance [LF$^+$90, A$^+$95a], but needs further confirmation. Evidence for the excited B states first emerged in analyses at LEP. The narrow $B_{u,d}^{**}$ states have been identified, although a separation of the states was not possible due to the detector resolution. As in the charm sector, the broad states have not been observed directly. Constrained fits to the spectra in the framework of HQET gave some hints without solving the problem of spin-orbit inversion. The mechanism of spin-orbit inversion is further discussed in section 3.7.1. Some theoretical papers predict the broad doublet to lie higher in mass than the narrow states, which is contrary to the usual understanding of hadron spectra. Beside the LEP experiments, the CDF Collaboration has presented results on $B_{u,d}^{**}$ production from a hadron collider by combining a charged primary pion to a partially reconstructed B meson. Unfortunately, the CDF measurement could also not provide any direct results on the broad states.

This thesis reports on a search for orbitally excited, $L = 1$, B-mesons in $Z^0 \to b\bar{b}$ events taken in the years 1994 to 2000 with the DELPHI detector at the LEP collider. The analyses presented represent the result of applying much improved and extended techniques of spectroscopy since the first DELPHI publication [A$^+$95b]. The tracking detectors have been optimised in the years 1995/1996 by better alignment procedures and the addition of completely new algorithms. All LEP I data taken from 1992-1995 have been reprocessed with these new algorithms, and this has led to huge reconstruction improvements, especially in dense jets. As an example, the signal of exclusively reconstructed D^* mesons into $D^0\pi^+$ with $D^0 \to K\pi\pi\pi$ was increased by a factor 2.4. A major improvement has occured in the area of particle identification, especially for hadron identification, where a neural network approach has been implemented in the DELPHI software package MACRIB [AFM99]. Developments in the area of neural networks have led to much improved enrichment of the excited B states. The B^{**} neural networks identify on a track-by-track basis the pion or the kaon originating from the B^{**} decay, suppressing background and keeping signal events in an efficient way. Further background suppression and isospin measurement are feasible with the application of the Charge Correlation Network taking into account the correlation of the signal track charge with the flavour of the b quark within its hadronic decay partner. The detector resolution plays a crucial role in the determination of the mass splittings between the states. To improve on this area, a further application of neural networks has been applied to reconstruct the underlying Q-value. The corresponding

network gives a correction on existing measurements of the Q-value in the form of a probability density function, thus not only giving the most probable value for the true quantity but also providing a measurement error, and this on a track-by-track basis. This allows an indivual selection of well measured tracks to be involved in the analysis. Many of these algorithms are part of the BSAURUS program package, a general tool for the inclusive study of b-hadrons with the DELPHI detector. Beside the above mentioned techniques, BSAURUS provides a variety of measurements essential for the presented analyses. Just to mention one, the network based b-hadron type identification allows to combine charged tracks with the oppositely charged $B^{\pm}$ meson, without having to cope with the backround consisting of B^0 and B_s mesons as well as of b-baryons.

The second chapter is devoted to the introduction of the theoretical framework and its predictions concerning the properties of excited B mesons. The quantum numbers and the decay modes of the states are derived, using the ideas of HQET. In the following chapter, the general concept of hadron spectroscopy as well as the specific approach taken for this thesis are explained. The fourth chapter gives an outline of the DELPHI detector and of the LEP collider, at which the detector is installed. The key point of the performed study is the usage of neural network techniques, motivating a separate chapter on their conceptual structure, their effective training and the interpretation of their results. Before coming to describe the analyses, the tools which have been mentioned briefly here, are explained in detail. Although the analysis results are the main message of this write-up, the tools developed to achieve them represent the main exertions in the day-to-day work. The analysis method is finally described in section 7 and the conclusions that can be drawn in section 8.

Chapter 3

Theoretical Background

This chapter provides the theoretical background needed for the understanding of the presented analyses. It is structured in three main sections: the creation, the properties and the decay of excited B mesons. The creation is described over the whole process of e^+e^- annihilation, the perturbative QCD phase and the confinement phase. The properties of B^{**} mesons are developed in terms of the Heavy Quark Effective Theory. Their decay properties, leading to the experimentally accessible quantities, are discussed thereafter.

3.1 Electron-Positron Annihilation into Hadrons

B^{**} mesons are produced at LEP at the end of a long chain of interactions characterized by different forces (see figure 3.1). The production chain can be subdivided into 4 time ordered phases: the electroweak phase of e^+e^- annihilation, the perturbative phase of the short-distance strong interaction of the produced quark-antiquark pair, the fragmentation phase and finally the decay into final state particles which are observable in the detector.

3.1.1 The Electroweak Phase

The first phase in e^+e^- annihilation is governed by electroweak interactions [Cab63]. The process $e^+e^- \to q\bar{q}$ proceeds to first order (Born) approximation by the exchange of a γ or a Z^0 boson. The fermion couplings to the Z^0 can be expressed in terms of a vector and axial-vector part. The total cross-section for $e^+e^- \to f\bar{f}$ at LEP is completely dominated by the presence of the Z^0 resonance. The situation is illustrated by figure 3.2.

The total width of the Z^0 resonance (≈ 2.495 GeV) is the sum of the partial widths of all produced fermions. The partial decay width for the channel $Z^0 \to q\bar{q}$ in lowest order and ignoring the quark mass is given by

$$\Gamma_{q\bar{q}} = \frac{G_F m_Z^3}{8\sqrt{2}\pi}(v_q^2 + a_q^2) \tag{3.1}$$

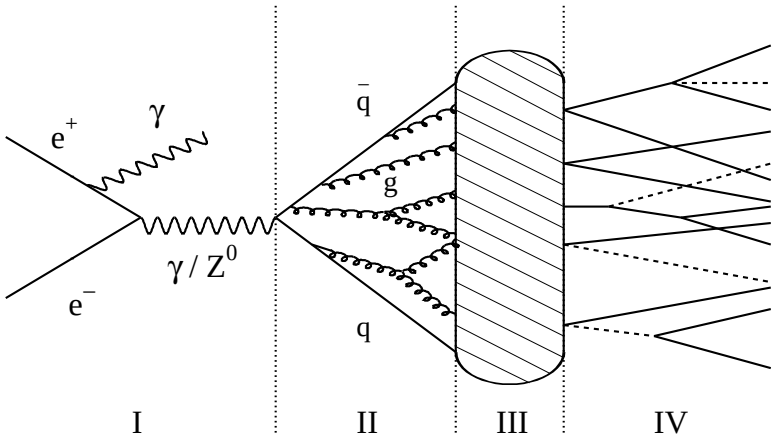

Figure 3.1: *The four phases of e^+e^- annihilation: I) electroweak phase with initial state radiation, II) perturbative QCD phase with gluon radiation off the leading quarks, III) confinement phase where colour singlet hadrons emerge and IV) decay into final state particles.*

where the vector and axial-vector coupling of a quark is

$$v_q = I_3^W(q) - 2Q_q^{em}\sin^2\theta_W \tag{3.2}$$

$$a_q = I_3^W(q). \tag{3.3}$$

This results in different partial widths for up-type ($I_3^W = 1/2, Q^{em}_{\text{up}} = 2/3$) and down-type quarks ($I_3^W = -1/2, Q^{em}_{\text{down}} = -1/3$). The sum of the hadronic partial widths $\Gamma(\text{had})$ is 1.7444 ± 0.0020 GeV resulting in the branching ratio $Br(Z^0 \to q\bar{q}) \approx 69.91\%$. The ratio R_b, which is of particular interest for this analysis, is

$$R_b = \frac{\Gamma_{b\bar{b}}}{\Gamma(\text{had})} = 0.21664 \pm 0.00068 \tag{3.4}$$

which is in good agreement with the Standard Model prediction of 0.21569 ± 0.00016 [H$^+$02]. Overall, only 15.15% of the Z^0 produce a $b\bar{b}$ pair. The analyses presented in this thesis consider only events of this type.

3.1.2 The Perturbative Phase

Starting from the produced quark-antiquark-pair, a configuration of colour charged partons (quarks and gluons) is created by perturbative QCD pro-

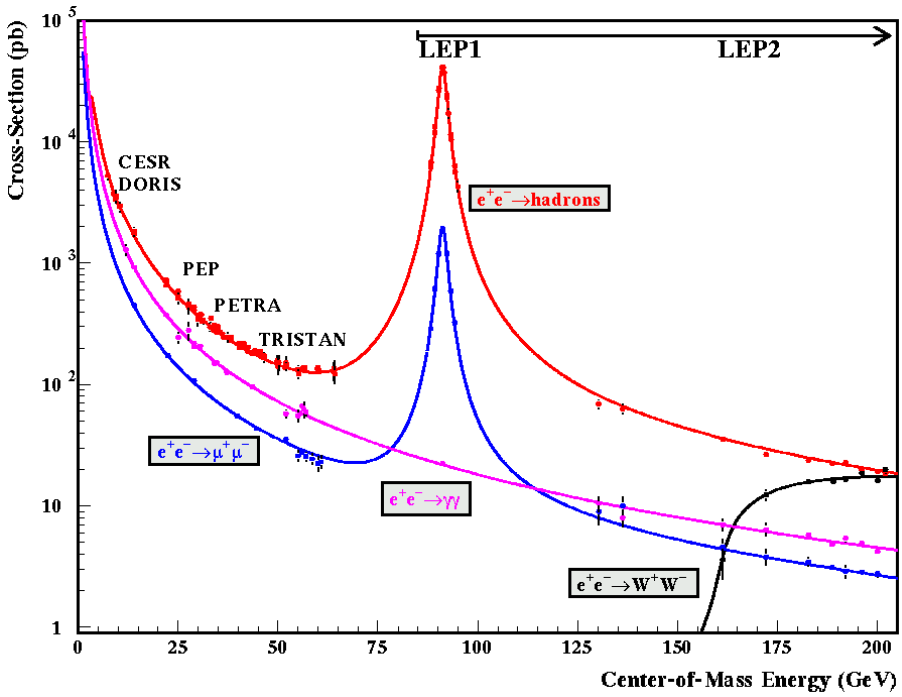

Figure 3.2: *Hadron production in e^+e^- annihilation at LEP energies. The curves show the expected $1/s$ dependence with the Z^0 pole for fermions which couple to the Z.*

cesses. Up to five created partons, the exact matrix elements can be used (Next-to-Next-to-Leading Order [MUW02]). These matrix elements are the summed contributions of thousands of QCD-Feynman-Diagrams per produced parton. The Matrix Elements approach was exclusively suitable for low energy experiments (see figure 3.2) where only up to 4 final state hadrons were produced. At the energies of LEP, a large amount of particles emerge from the e^+e^- interaction making the use of exact matrix elements impossible, since computational time would increase to infinity.

Instead, a model is introduced that is capable of simulating the observed data. This alternative comes from the use of the leading logarithm approximation. In this approach, only the leading terms of the perturbative expansion are kept. The parton shower model (LEPPS) is based on the picture of a time ordered cascade of subsequent splitting processes of the partons $q \to qg, g \to q\bar{q}$ and $g \to gg$. The evolution of the shower via parton splitting is stopped when an energy scale $Q_0 \approx 1$ GeV is reached. The exact number used for the simulation of the 1994 and 1995 data is given in table 8.3. The transition to colour neutral hadrons is described in the next section.

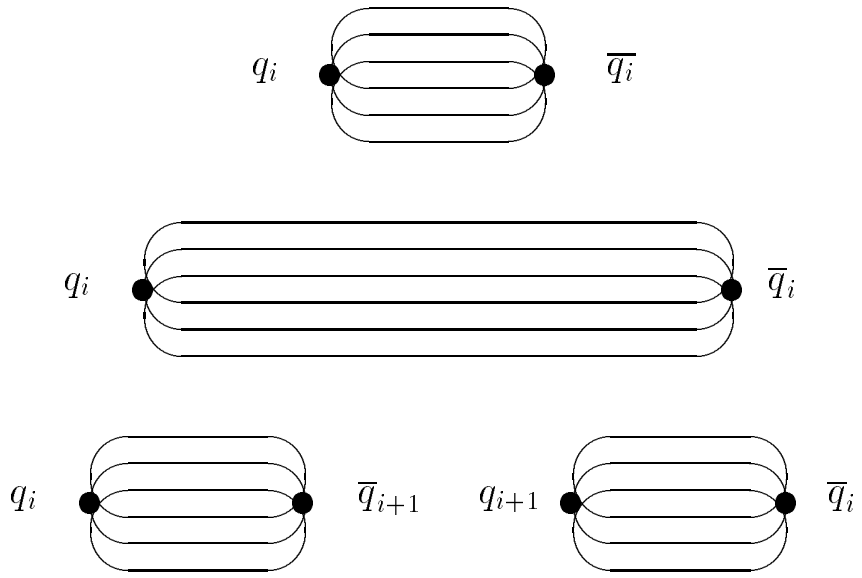

Figure 3.3: *Schematic view of the string fragmentation algorithm. The self interaction of the colour field leads to a linear field of constant energy density called the string. The string breaks and a meson is formed. The rest of the string is left as an exact copy of the original scaled down by the energy taken by the meson.*

3.1.3 The Fragmentation Phase

Fragmentation describes the formation of colour neutral hadrons from colour charged quarks and gluons. The process is not perturbatively computable since $\alpha_s \approx 1$ for low Q^2 and thus has to be phenomenologically modelled. To describe the data, a set of parameters has to be adjusted. A successful model is provided by the string fragmentation [Mor89, AGIS83].

It starts with a pair of massless quarks and antiquarks moving in opposite directions with $v = c$. The self coupling of the gluons leads to a linear field configuration with the typical size of 1 fm, which is spanned between the quarks. This object is called string. The linear energy density in the colour field is a constant $\kappa \backsimeq 1$ GeV/fm leading to a linearly growing potential between the quarks. As soon as the energy stored in the field exceeds the quark pair creation threshold, a new meson is formed leaving the rest of the string as an exact copy of the original, scaled down by the energy taken by the meson. Figure 3.3 schematically shows this procedure.

The flavour content of the new hadrons depends on the probability of creating a quark-antiquark-pair within the thin linear colour flux tube. The probability for such a pair creation process can be calculated by a simple quantum mechanical approach treating the effect as a tunneling phenomenon. Assuming constant linear energy density κ, the tunneling proba-

bility is given by

$$\exp\left(-\frac{\pi m^2}{\kappa}\right)\exp\left(-\frac{\pi p_T^2}{\kappa}\right)=\exp\left(-\frac{\pi m_T^2}{\kappa}\right) \tag{3.5}$$

where m_T is the common transverse mass of the produced quark-antiquark pair. The p_T distribution of mesons is assumed to be Gaussian with zero mean and standard deviation σ_{p_T} (see table 8.3). The formula implies a suppression of heavy quark production

$$u:d:s:c:b\approx 1:1:0.3:10^{-11}:10^{-100}. \tag{3.6}$$

This means that c and b quarks are practically never produced by the soft hadronization mechanism. Since the light quark masses are only known with large errors, the s quark suppression $u:d:s=1:1:\gamma_s$ is a free parameter of the model (see table 8.3).

What remains to be determined is the energy and longitudinal momentum of the hadron. In fact, only one variable can be selected independently, since the momentum of the hadron is constrained by the already determined hadron transverse mass m_T

$$E^2-p_z^2=m_T^2=m^2+p_x^2+p_y^2=m^2+p_T^2. \tag{3.7}$$

The fraction of energy and momentum z, which is given to the meson, is distributed according to the longitudinal fragmentation function $f(z)$ with z defined as

$$z=\frac{(E+p_\parallel)_{hadron}}{(E+p)_{q_0}} \tag{3.8}$$

for 2-jet events, where $p_\parallel$ is the momentum along the direction of motion of the original quark q_0 and $(E+p)_{q_0}$ is the total available energy. Requiring that the fragmentation process as a whole should look the same, irrespectively of whether the iterative procedure is performed by the q or the $\bar{q}$ end, the choice of the fragmentation function is essentially unique: the *symmetric Lund fragmentation function*

$$f(z)=\frac{z^{a_\alpha}}{z}(\frac{1-z}{z})^{a_\beta}\exp(-\frac{bm_\perp^2}{z}) \tag{3.9}$$

where a_α and a_β are separate parameters for each flavour. In addition, there is the universal parameter b. This fragmentation model has been used for u, d and s quarks in the DELPHI Monte Carlo, with a_α and a_β set to the values listed in table 8.3. For the fragmentation of the c and the b quark, the Peterson fragmentation function

$$f(z)=\frac{1}{z(1-1/z-\epsilon_q/(1-z))^2} \tag{3.10}$$

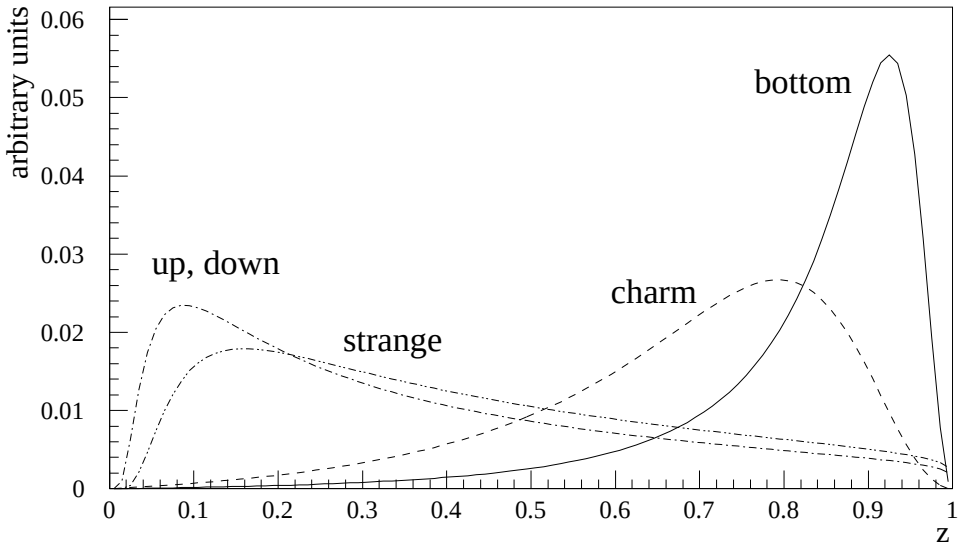

Figure 3.4: *The longitudinal fragmentation functions for the different quark flavours. The heavy flavours are described by the Peterson function with $\eta_b = 0.006$ and $\eta_b = 0.054$. For the light flavours the Lund symmetric function is plotted with the values $a = 0.18$, $b = 0.34$, $\sigma_{p_T} = 395 MeV/c$ and the masses of pion and kaon as typical representations for a light and strange meson.*

has been used. The Peterson function has a flavour dependent parameter ϵ_q. It is the squared ratio of the masses of heavy to light quarks. However, since the light quark masses are not precisely known, in practice the values of $\epsilon_{c,b}$ are evaluated from data. Once the parameter is known for one heavy flavour, the appropriate value for the other can be obtained from the quark mass relation $m_c^2 \epsilon_c = m_b^2 \epsilon_b$. As shown in figure 3.4, the fragmentation function of the heavy quarks is peaked at high values of z. Due to this fact, hadrons containing the primary c or b quark have a large fraction of the beam energy. Also shown is the distribution of the symmetric Lund function for the light quarks. The actual parameters used for simulating the 1994 and 1995 data is summarized in table 8.3.

In the string model, gluons are treated as transversal excitations of the string. Due to their colour composition, they are attached to the end of two strings. In a $q\bar{q}$ configuration, a string is connected between the quark and the gluon and a second is connected from the gluon to the antiquark.

The creation of a diquark in the fragmentation phase leads to baryon production. For this case, the same statement is true about the suppression factors for heavy flavours. In order to conserve quantum numbers, a second diquark has to be produced in association with the first diquark. It can either be produced immediately next to the first diquark or with some mesons in

between. The latter case is known as pop-corn effect.

3.1.4 Hadron Decay

In the last phase, the hadrons produced in the fragmentation phase decay into stable particles. These decays are simulated in Monte-Carlo programs referring to experimentally measured branching ratios. Especially within the domain of bottom hadrons, a number of states for which no or only a few experimental data exist, is produced. The modelling of their production and decay is inspired by predictions and observations in analogous systems, quantum number conservation and simple spin and state counting arguments.

3.2 Quantum Chromodynamics

The production of hadrons at LEP has been described in the last section. The characteristics of these hadrons (mesons and baryons) are described by the theory of strong interactions. The quantum field theory of the strong interaction is called Quantum Chromodynamics (QCD). It explains why quarks are always bound in colour singlet hadrons and the reason why, at large energies, quarks asymptotically become free.

Quarks and gluons are the only strongly interacting particles. They carry strong charge mediated by the gluons, the quanta of QCD, which are massless vector bosons with spin 1, like photons. The main difference to photons is that gluons themselves carry charge resulting in gluon self interaction. The charge of QCD are the colours represented by *red*, *green* and *blue*. The number of colours comes from observations in data. There are 8 gluons which carry net colour and one colour singlet gluon which does not take part in strong interactions. Quarks in hadrons are bound together by gluon exchange.

3.2.1 Asymptotic Freedom and Confinement

QCD is, on the one hand, a renormalizable theory and therefore treatable by perturbative methods. Perturbation theory is appropriate for QCD at high energy or high momentum transfers. This characteristic of QCD is called *asymptotic freedom*. On the other hand, QCD becomes non-perturbative at low energies as a consequence of its non-abelian nature.

The one loop approximation of the momentum dependence of the strong fine-structure constant α_s can be brought into the simple form

$$\alpha_s(Q^2) = \frac{4\pi}{\beta_0 \log \frac{Q^2}{\Lambda^2_{QCD}}} \tag{3.11}$$

Quantum Number	u	d	s	c	b	t
Q Electric Charge	$+\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$+\frac{2}{3}$	$-\frac{1}{3}$	$+\frac{2}{3}$
P Parity	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$
I_3 Isospin (3. component)	$+\frac{1}{2}$	$-\frac{1}{2}$	0	0	0	0
S Strangeness	0	0	-1	0	0	0
C Charm	0	0	0	$+1$	0	0
B Bottomness	0	0	0	0	-1	0
T Topness	0	0	0	0	0	$+1$

Table 3.1: *Quantum numbers of quarks. The convention is that the flavour of a quark (described by I_3, S, C, B and T) has the same sign as its charge. Antiquarks have by definition opposite quantum numbers since quarks are fermions.*

where Λ_{QCD} is the asymptotic scale parameter of QCD. This equation illustrates the asymptotic freedom property of QCD: $\alpha_s \to 0$ as $Q^2 \to \infty$ and shows that QCD becomes strongly coupled at $Q^2 \sim \Lambda_{QCD}$. In the low energy region, quarks are confined into hadrons.

3.2.2 Bound States

Hadrons, i.e. bound states of quarks, are colour-singlet combinations of quarks, antiquarks and gluons. Only two types of bound states have been experimentally observed: *mesons* and *baryons*. The decay scheme of hadrons into lighter hadrons has led Gell-Mann and Zweig to the quark model. The strength of particular decay channels or even the non-observation of certain channels could be explained by assigning different quantum numbers to the quarks and by demanding that these quantum numbers are conserved in strong interactions. The quantum numbers of quarks are given in table 3.1.

Mesons

Mesons are bosons with integer spin and are bound states of a quark q and an antiquark $\bar{q}$, where the flavours of the quark and antiquark may be different. The intrinsic parity of a meson is $-1 = (+1)\cdot(-1)$ since the quark has parity $+1$ and the antiquark has parity -1 and parity is a multiplicative quantum number. The orbital parity of the quark-antiquark system is determined by the angular momentum L to be $(-1)^L$ which follows from the symmetry of spherical functions. The total parity of mesons is thus $P = (-1)^{L+1}$. A state $q_i\bar{q}_i$ of a quark and its own antiquark is also an eigenstate of charge conjugation C, with the C-Parity $\eta_C = (-1)^{L+S}$, where the total spin S is 0 or 1. The $L = 0$ states are the pseudoscalars ($J^{PC} = 0^{-+}$, $S = 0$) and the vectors ($J^{PC} = 1^{-+}$, $S = 1$). The properties of the $L = 1$ states are treated in detail in section 3.5.

The work of this thesis is devoted to B meson spectroscopy of the excited states with $L = 1$. A discussion of the present experimental status of B spectroscopy as well as a short review of the results in the D meson sector will be given in chapter 4. The theoretical background of B meson ground states and excited states is discussed in detail in section 3.5.

Baryons

Baryons are fermions with half-integer spin and are 3-quark states (qqq) or 3-anti-quark states ($\bar{q}\bar{q}\bar{q}$). Each of these states is a SU(3) colour singlet, a completely antisymmetric state of the three possible colours. All of the 'ordinary' baryons containing u,d and s quarks have been observed in experiments. Most of the charmed baryons have been seen. Bottom baryons are not yet established apart from the Λ_b. There is evidence of other b baryons like the Ξ_b baryon [BLMÅ02] but the experimental results have not yet been published.

3.3 Production Rates of b Hadrons

The spectroscopy of excited B mesons relies not only on the exact reconstruction of the decay product momenta but requires a good knowledge of b hadron production rates. The measurement of the B^{**} rates depends on the rates of mesons produced in the fragmentation phase, on the rate of picking up the corresponding quark (a u quark for the B^{+**}, a d quark for the B^{0**} and an s quark for the B_s^{0**}) and on the rate of producing excited state mesons.

3.3.1 Primary Hadron Production Rates

The high centre-of-mass energy of LEP $\sqrt{s} \gg 2m_b$ results in a large Lorentz boost of the produced hadrons. This, together with the hard b fragmentation function, results in a measurable decay length of the weakly decaying b hadrons. The discussion of production rates thus has to make a distinction between primarily produced hadrons and hadrons decaying via the weak interaction. The fragmentation models discussed above can, supplied with appropriate parameters, predict rates of primary b hadrons. The parameters used for the DELPHI Monte Carlo simulation are summarized in table 8.3 and are schematically shown in figure 3.5. These numbers primarily refer to the quark content of the hadrons regardless of the level of excitation. The rates are consistent with the numbers given in the PDG for the

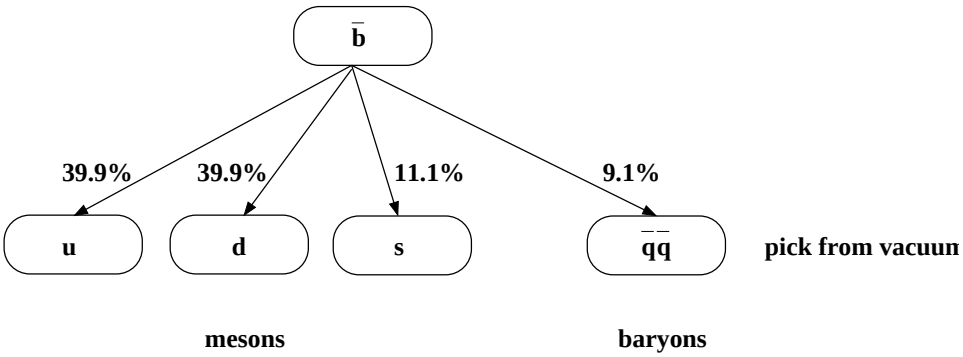

Figure 3.5: *Rates of primarily produced b hadrons. The rates are the result of the used parameters in the DELPHI Monte Carlo simulation (see table 8.3).*

$B^{\pm}/B^0/B_s^0/b$-baryon admixture [H$^+$02]:

$$\begin{aligned} B^+ &\quad (38.8 \pm 1.3)\% \\ B^0 &\quad (38.8 \pm 1.3)\% \\ B_s^0 &\quad (10.6 \pm 1.3)\% \\ b-\text{baryon} &\quad (11.8 \pm 2.0)\% \end{aligned} \tag{3.12}$$

Before coming to the production rates of excited B mesons, their properties are discussed in the light of the heavy flavour and heavy spin symmetries. These symmetry principles allow to draw a simplified picture of the states and the prediction of the rates becomes the relatively straightforward quantum mechanical approach of spin or state counting.

3.4 Symmetries in QCD

The non-abelian nature of QCD makes predictions using analytic methods impossible. Instead, predictions are based on symmetries of QCD: the isospin, the chiral and the heavy quark symmetry (HQS).

3.4.1 Isospin Symmetry

Isospin symmetry is understood as an approximate symmetry which arises because the light quark mass difference $m_d - m_u$ is much smaller than the mass associated with confinement, which is set by the scale Λ_{QCD}. The predictions based on isospin symmetry would, in a world with only strong interactions, be exact in the limit $(m_d - m_u) \to 0$. Corrections to this limit can be studied in an expansion in the small parameter $(m_d - m_u)/\Lambda_{QCD}$. $SU(3)$ flavour symmetry is similar but the corrections are larger since $(m_s - m_u)/\Lambda_{QCD}$ is not small.

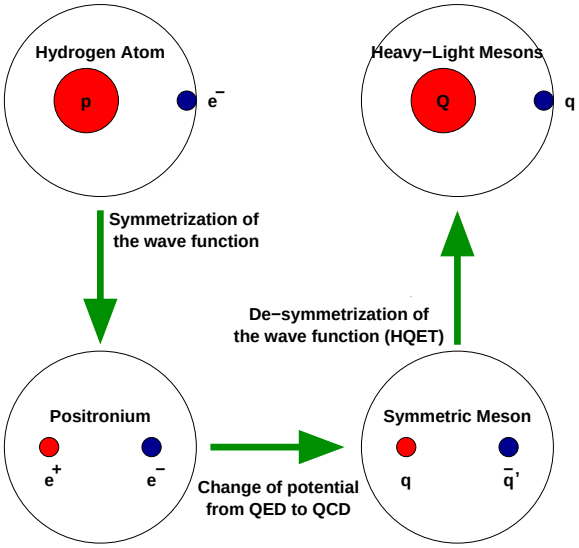

Figure 3.6: *From the Hydrogen atom to B mesons. The proton can be taken as a static source of electrical charge since it is much heavier than the electron . If the proton is replaced by a positron, both particles effect each other equally. The same is true for a symmetric meson with a different kind of potential. A heavy-light meson is analogous to the H-atom since the spin of the heavy quark s_Q and the total angular momentum of the light degrees of freedom j_q are conserved separately.*

3.4.2 Chiral Symmetry

Chiral symmetry $SU(2)_L \times SU(2)_R$ arises in QCD because both m_d and m_u are small compared to Λ_{QCD}. It is associated with the separate conservation of vector and axial vector currents. Although spontaneously broken in nature, the existence of this underlying symmetry allows the expansion of chiral perturbation theory in which many low-energy properties of QCD are related to a few reduced matrix elements. If the strange quark is also treated as small compared to Λ_{QCD}, the chiral symmetry group becomes $SU(3)_L \times SU(3)_R$.

3.4.3 Heavy Quark Symmetry

Consider a bound state of a heavy-light quark-antiquark $Q\bar{q}$. In the limit $m_Q \to \infty$, the heavy quark can be taken as a static source of a colour field in which the light quark q moves. The situation is analogous to the hydrogen atom (figure 3.6). In the limit $m_Q \to \infty$, the Hamiltonian for the light degree of freedom, in analogy with the H-Atom, can be written in the form:

$$-\frac{\mathbf{p}^2}{2m_q}\Psi(\mathbf{p}) = \int \frac{d^3q}{(2\pi)^2} V(\mathbf{p},\mathbf{q};m_q)\Psi(\mathbf{q}) \tag{3.13}$$

The effective heavy quark theory provides a framework to take into account $1/m_Q$ corrections due to the finite mass of the heavy quark. Starting from the Dirac equation and separating the heavy meson field into a heavy and light component quark, the Lagrangian for the heavy quark field is given by

$$\mathcal{L}_{\text{eff}} = \bar{Q}\, ivD\, Q + \frac{1}{2m}\bar{Q}\left[D^2 + (ivD)^2\right] Q + \frac{1}{4m}\bar{Q}\, g_s \sigma^{\mu\nu} G_{\mu\nu}\, Q \tag{3.14}$$

where the first term is related to the velocity of the heavy quark, the second term is its kinetic energy and the third corresponds to its colour magnetic moment. The second and third terms violate flavour and spin symmetry for a finite heavy quark mass. In the infinite mass limit only the first term is non-vanishing:

$$\mathcal{L}_{\text{eff}} = \bar{Q}\, ivD\, Q\ . \tag{3.15}$$

The residual motion of the heavy quark can be taken into account by adding the effective Hamiltonian to the Hamiltonian for the light quark. In the infinite mass limit, the effective Lagrangian $\mathcal{L}_{\text{eff}}$ has additional symmetries not present in the full QCD Lagrangian.

The *spin symmetry* (HSS) of the heavy quark that arises is reflected in the fact that the first term in the effective Lagrangian makes no reference to the Dirac structure at all which can couple to the spin degrees of Q. Thus, the masses of vector and pseudoscalar mesons are degenerate in this limit. This degeneracy is lifted by the third term in the Lagrangian. For example, for the ground state (1^-) and (0^-) mesons, the mass difference, i.e. the hyperfine splitting, scales like $1/m_Q$.

The second symmetry of the Lagrangian in the limit $m_Q \to \infty$ arises when distinct flavours, e.g. b and c, are introduced. Since the first term makes no reference to masses, and since mass is the only property which can distinguish between quarks of different flavours in QCD, the effective theory has a symmetry under which $b \leftrightarrow c$. This symmetry does not depend on $m_c = m_b$ but on $m_i \gg \Lambda_{QCD}$. *Heavy Flavour Symmetry* (HFS) thus holds between heavy quark fields of the same velocity and not of the same momentum. Flavour symmetry together with the spin symmetry gives rise to a $SU(4)$ symmetry for the system $(b(v)\, c(v))$.

Applications of HQET

Though HQET provides a framework in which the properties of excited B mesons can be derived, the validity for its application to heavy quarks is tested in several other fields. The most important application of HQET

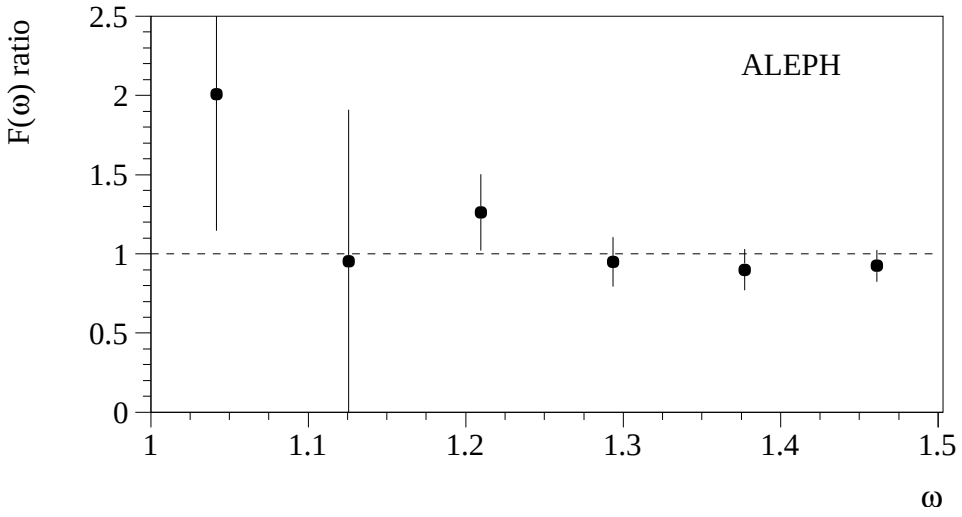

Figure 3.7: *Ratio of the two form factors $\mathcal{G}(w)$ and $\mathcal{F}(w)$, which is predicted to be unity in the heavy quark limit, testing Heavy Quark Symmetry [B+ 97].*

concerns the description of exclusive semileptonic decays based on the quark transition $b \to c\, l\, \bar{\nu}$. This is where the theory is well tested, and theoretical uncertainties are best understood. The most important result is a precision determination of the CKM parameter $|V_{cb}|$ [Neu94, A+01b].

In general, the decays $\bar{B} \to D\, l\, \bar{\nu}$ and $\bar{B} \to D^*\, l\, \bar{\nu}$ are described by the form factors: $\mathcal{G}(w)$ for the first process, and $\mathcal{F}(w)$ for the second one. In the heavy quark limit, $\mathcal{G}(w)$ and $\mathcal{F}(w)$ become equal to the "universal" function $\xi(w)$. The universality of the function $\xi(w)$ can be tested by measuring the ratio $\mathcal{G}(w)/\mathcal{F}(w)$ as a function of w. The ALEPH data for this ratio are shown in figure 3.7. Within errors, the data are compatible with a universal form factor. At large recoil, where the experimental errors are the smallest, this provides a test of heavy quark symmetry at the level of $10 - 15\%$.

3.5 Properties of B Mesons

Since, accoriding to HQET, the spin of the heavy quark decouples, we should find states occurring in doublets corresponding to the two possible orientations of the heavy quark spin. We can classify states using the heavy quark spin s_Q and the total spin j_Q of the light quark, which are both good quantum numbers. In this section, the quantum numbers of the ground state and of the excited B mesons are discussed.

For the discussion of mesonic states, the usual spectroscopic notation is used: $n^{2S+1}L_J$, where n is the radial quantum number[1], S the total spin, L

[1] note: here $n = n_r$. In atomic physics $n = n_r + l$.

the orbital angular momentum between the quarks and $\vec{J} = \vec{L} + \vec{S}$.

3.5.1 Ground State B Mesons

According to HQET, each energy level in the excitation spectrum of $(Q\bar{q})$ mesons is composed of a degenerate pair of states characterized by $\vec{j_q} = \vec{s}_q + \vec{L}$ and the total spin $\vec{J} = \vec{j_q} + \vec{s}_Q$, i.e. by $J = j_q \oplus \frac{1}{2}$. The ground state mesons with $L = 0$ correspond to $\vec{J} = \frac{1}{2} \oplus \frac{1}{2}$, with $J^P = 0^-$ for the pseudoscalar B mesons and $J^P = 1^-$ for the B^* vector mesons. In the Dirac notation, the ground states can be written as:

$$\begin{aligned} |B^*, +1\rangle &= |+1/2, +1/2\rangle, \\ |B^*, 0\rangle &= \frac{1}{\sqrt{2}}\left(|+1/2, -1/2\rangle + |-1/2, +1/2\rangle\right), \\ |B^*, -1\rangle &= |-1/2, -1/2\rangle, \\ |B, 0\rangle &= \frac{1}{\sqrt{2}}\left(|+1/2, -1/2\rangle - |-1/2, +1/2\rangle\right), \end{aligned} \tag{3.16}$$

where the first $\pm 1/2$ is the 3rd component of the spin of the heavy quark and the second $\pm 1/2$ is the 3rd spin component of the light degrees of freedom.

Production Rate of Ground State B Mesons

It can be seen that the pseudoscalar B meson belongs to a singlet, while the vector B^* meson belongs to a triplet. From a simple *spin counting* argument, one would expect a production ratio of $B^* : B$ mesons of $3 : 1$ in e^+e^- collisions. This simple argument needs slight corrections due to the production of higher excited states. This will be discussed in the next section. This expectation is in accord with LEP measurements $\sigma_{B^*}/(\sigma_{B^*} + \sigma_B) = 0.75 \pm 0.04$ [H$^+$02].

3.5.2 B Mesons with Orbital Excitation

The spin j_q of the light quark can, for $L = 1$, take two values $j_q = \frac{1}{2} \oplus 1$: $j_q = \frac{1}{2}$ and $j_q = \frac{3}{2}$. The addition of the heavy quark spin induces a splitting of these states into two doublets with the quantum numbers J^P given in table 3.2. The official notation is given in the third column. All four states are commonly labelled as B^{**}. In order to describe whether the $\bar{b}$ quark is accompanied by a u, d or an s quark, the notation is extended by the subscript u, d or s. This thesis focuses on the search for and on the analysis of these states.

Production Rate of Excited B Mesons

A primarily produced $\bar{b}$ quark will pick up a quark q from the vacuum 90.9% of the time to form a meson, whereas in 9.1% of the time a baryon

j_{light}	J^P	B^{**} state	Decay Mode	Transition	Prod. Rate
1/2	0^+	B_0^*	$B\pi$	S-wave	1
1/2	1^+	B_1	$B^*\pi$	S-wave	3
3/2	1^+	B_1	$B^*\pi$	D-wave	3
3/2	2^+	B_2^*	$B^*\pi, B\pi$	D-wave	5

Table 3.2: *The four B^{**} states and their expected spin/parity assignments and decay properties. Also listed are the relative production rates of the states based on (2J+1) spin counting.*

is produced. From the D meson sector it is expected that roughly 30% of the mesons will be produced with $L = 1$ and 70% with $L = 0$. The validity of the spin counting approach for the production rate of the ground state mesons leads to the assumption of the relative rates $1 : 3 : 3 : 5$ for the excited states $B_0^* : B_1 : B_1 : B_2^*$ respectively (table 3.2).

Influence of B^{**} Production on the Ratio B^*/B

As discussed in section 3.5.1, the production ratio of B^* mesons per primary ground state mesons was measured by the LEP experiments to be: $\sigma_{B^*}/(\sigma_{B^*} + \sigma_B) = 0.75 \pm 0.04$ [H$^+$02]. In the absence of B^{**} production, this ratio is equal to the parameter $V/(V + P)$ in the JETSET fragmentation model, where V and P are the production rates of primary vector and pseudoscalar B mesons. However, production of a sizeable amount of B^{**} mesons can alter the ratio of B^* to B mesons depending on the relative production rates and branching fractions into $B^*\pi$ and $B\pi$ of the four individual B^{**} spin-parity states. Different assumptions for the production ratios and the unknown 2^+ decay branching ratio into $B^*\pi$ may change the effective branching fraction of B^{**} to $B^*\pi$. This implies that

$$\frac{V}{V+P} = \frac{1}{1-f^{**}} \left(\frac{\sigma_{B^*}}{\sigma_{B^*} + \sigma_B} - f^{**} \cdot \mathrm{Br}(B^{**} \to B^*) \right) \tag{3.17}$$

where f^{**} denotes the production ratio of B^{**} mesons per primarily produced meson. The measurement of f^{**} is one of the main goals of this thesis.

Influence of B_s^{**} Production for B Meson Rates

According to isospin rules, B_u^{**} and B_d^{**} will decay in 67% of the time with the emission of a charged pion and in 33% with the emission of a neutral pion, if the $B^{(*)}\rho$ channel is neglected. In this analysis only the charged pion decay channel is investigated since the detection of the π^0 needs the conversion of both decay photons, a process with very low efficiency in the DELPHI

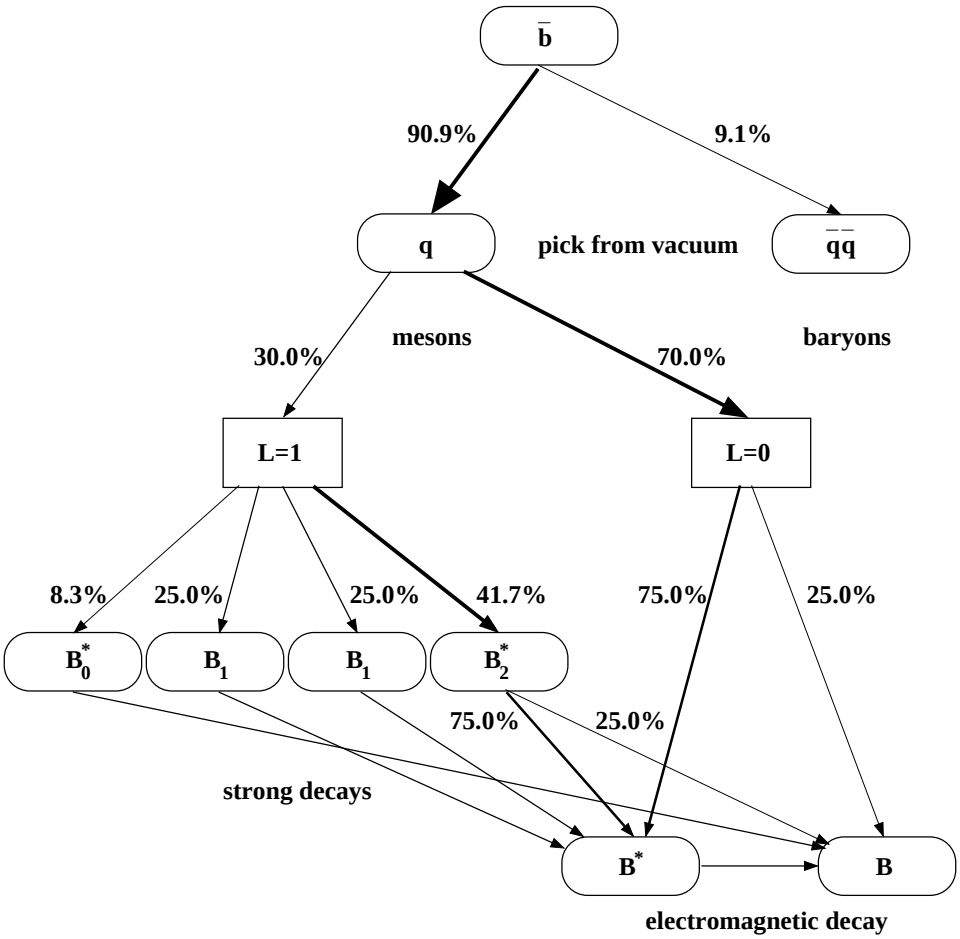

Figure 3.8: *Production rate of excited B mesons, as predicted by HQET. Deviations from these numbers occur due to mass differences between the states resulting in a different dynamical suppression of the decay modes.*

detector. The decay $B_s^{**} \to B_s^{(*)}\pi^0$ is forbidden by isospin conservation. Instead, B_s^{**} mesons decay by the emission of a kaon ($50\%\, K^+$, $50\%\, K^0$).

Assuming the usual strangeness suppression factor $u : d : s = 3 : 3 : 1$ for the production of B_s mesons in b events, fragmentation leads to a production ratio of $\sigma_{B_s}/(\sigma_B + \sigma_{B_s}) = 14.3\%$. This production ratio has to be corrected if sizeable amounts of B_s^{**} mesons are produced, since they do not decay into B_s mesons. One obtains

$$\frac{\sigma_{B_s}}{\sigma_B + \sigma_{B_s}} = 14.3\% \cdot (1 - f_s^{**}), \tag{3.18}$$

where f_s^{**} is the production ratio of B_s^{**} mesons per primarily produced B state.

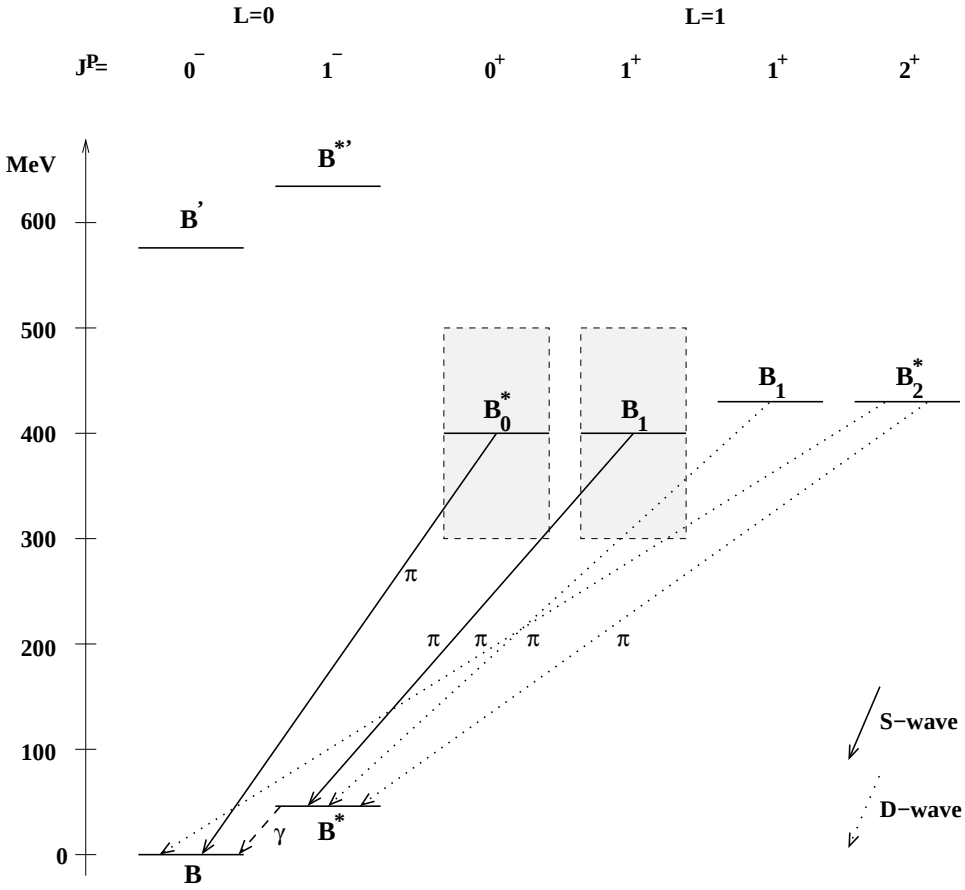

Figure 3.9: *The spectrum of excited B mesons relative to the ground state pseudoscalar B with $m_B = 5.2789$ GeV and the corresponding transition lines. The transition lines of the radially excited B mesons were omitted for clarity. For B_s^{**} mesons, the decay pions have to be replaced by kaons.*

3.6 Decay of B Mesons

The decay of B mesons is characterized by electromagnetic and strong decays to the ground state B mesons and by the consecutive weak decay of the (0^-) state to lighter hadrons. Since the decay modes of the excited states are dependent on the masses themselves, the following discussion can just give an approximate picture of the observable channels.

3.6.1 Decay of Ground State B Mesons

The 1^- partner B^* of the hyperfine doublet decays electromagnetically to the 0^- state since the predicted mass difference is of the order 40 MeV (e.g. [GEF02]), which is much smaller than the mass of the lightest hadron

$m_\pi \approx 140$ MeV. This is different in the D sector, where the shift between the ground states is slightly larger than the pion mass.

3.6.2 Decay of Excited B Mesons

Orbitally excited B mesons will decay due to the strong interaction. The expected main decay modes of $B^{**}_{u,d}$ mesons are $B\pi$ and $B^*\pi$. If the B^{**}_s meson mass is above the $B^{(*)}K$ threshold, this channel will be the dominant decay mode since the decay into $B_s\pi^0$ is forbidden by isospin conservation. For the case the B^{**}_s mass is below the $B^{(*)}K$ threshold, the decay would occur into the isospin forbidden $B_s\pi^0$ state or would decay electromagnetically. The states of the first doublet ($j_q = 3/2$) are narrow ($\Gamma \approx 10$ MeV) compared to the typical strong decay width, because only the $L = 2$ (d-wave) decays are allowed. This is due to the conservation of angular momentum and parity for the 2^+ state and for the 1^+ partner. The members of the second doublet ($j_q = 1/2$) are expected to be broad (a few hundred MeV), because only $L = 0$ (s-wave) decays are allowed. HQET predicts the two 1^+ eigenstates to be a $45°$ mixture of the states with $S = 0$ and $S = 1$. This results in the property that the 1^+ ($j_q = 1/2$) state decays via s-wave, while the 1^+ ($j_q = 3/2$) state decays via d-wave. The situation is similar to the D and K meson sector. Spin-parity rules govern the possible decay modes for the different states:

$$\begin{aligned} \text{broad states (S-wave decay)} \quad & \begin{cases} B_0^* \to B\,\pi \\ B_1 \to B^*\,\pi \end{cases} \\ \text{narrow states (D-wave decay)} \quad & \begin{cases} B_1 \to B^*\,\pi \\ B_2^* \to B^{(*)}\,\pi \end{cases} \end{aligned} \tag{3.19}$$

where the symbol $B^{(*)}$ denotes a B or a B^* meson, and for B^{**}_s decays the pion has to be exchanged by a kaon. Figure 3.9 gives an overview of the decay properties of orbitally excited B mesons.

3.7 Predictions for Orbitally Excited B Mesons

The properties of $L = 1$ heavy-light mesons are predicted by numerous theoretical approaches. The masses of the excited states can be calculated by referring to the dynamical structure of the light quark, determining the properties of the heavy-light system almost entirely by its own. Deviations from these calculations originate in the finite mass of the heavy quark resulting in $1/m_Q$ corrections of the meson spectrum. Predictions of the hadronic decay widths of these states are far more difficult since they are dependent of the masses themselves. In the next section, the approaches available today are outlined and compared with each other as well as with the data existing at present (table 3.3). A comparison with the results achieved in this analysis is performed in chapter 8.

3.7.1 The Prediction of Masses

The simplest approach of mass prediction has been performed by Falk and Mehen [FM96]. It is based on heavy spin symmetry combined with heavy flavour symmetry relating the masses of the narrow orbitally excited B mesons to the meson masses in the charm sector. The splittings between excited doublets and the ground state should be independent of the heavy quark mass, whereas the splitting within the doublet scales like $1/m_Q$. If one defines the spin averaged mass splittings to be:

$$\begin{aligned} \bar{m}_B &= \frac{3}{4}m_{B^*} + \frac{1}{4}m_B \ , \\ \bar{m}_{B^{**}} &= \frac{5}{8}m_{B_2^*} + \frac{3}{8}m_{B_1} \ , \\ \Delta m_{B^{**}} &= m_{B_2^*} + m_{B_1} \ , \end{aligned} \tag{3.20}$$

and analogous relations for charm mesons, the heavy quark symmetries predict that:

$$\begin{aligned} \bar{m}_{B^{**}} - \bar{m}_B &= \bar{m}_{D^{**}} - \bar{m}_D \ , \\ \Delta m_{B^{**}} &= \frac{m_c}{m_b}\Delta m_{B^{**}} \ . \end{aligned} \tag{3.21}$$

With the assumption $m_c/m_b = 1/3$ and averaging over the charged and neutral charmed mesons, these relations yield for the narrow states [FM96]:

$$\begin{aligned} m(B_1) &= 5780 \text{ MeV/c}^2 \\ m(B_2^*) &= 5794 \text{ MeV/c}^2 \\ m(B_{1s}) &= 5886 \text{ MeV/c}^2 \\ m(B_{2s}^*) &= 5899 \text{ MeV/c}^2 \ . \end{aligned} \tag{3.22}$$

Compared with the data in table 3.3, one can observe that the measured masses are somewhat lower than expected, especially for non-strange mesons. The leading corrections to the predictions are of the order:

$$\delta \sim \Lambda_{QCD}^2 \left(\frac{1}{2m_c} - \frac{1}{2m_b} \right) \sim 40\text{MeV} \ , \tag{3.23}$$

where $\Lambda_{QCD} \sim 400$ MeV has been assumed. It becomes evident that, together with measurements in the charm system, heavy quark symmetry provides powerful tools for predictions in excited B systems. The simplicity of these considerations increases the confidence in the predictions.

A more sophisticated model to estimate the B spectrum, where heavy quark symmetry is supplemented by insights gleaned from potential models, has been developed in two papers by Eichten, Hill and Quigg [EHQ93, EHQ94]. Their concept is based on the mass predictions of heavy quark

Quantum Numbers		State	Experiment [MeV/c^2]	Theoretical Predictions [MeV/c^2]				
J_l^P	J^P		PDG 2002	F+M	E+H+Q	Isgur	E+G+F	Khan
$\frac{1}{2}^-$	0^-	B^0	5279.4 ± 0.5				5285	5296
	0^-	$B^\pm$	5279.0 ± 0.5				5285	
	0^-	B_s	5369.6 ± 2.4				5375	5385
	1^-	B^{*0}	5325.0 ± 0.6				5324	5319
	1^-	B_s^*	5416.6 ± 3.5				5412	5412
$\frac{1}{2}^+$	0^+	B_0^*				5870	5738	5670
	0^+	B_{s0}^*					5841	5742
	1^+	B_1				5875	5757	
	1^+	B_{s1}					5859	
$\frac{3}{2}^+$	1^+	B_1	5698 ± 8	5780	5755	5700	5719	
	1^+	B_{s1}	5853 ± 15	5886	5834		5831	
	2^+	B_2^*	5698 ± 8	5794	5767	5715	5733	5822
	2^+	B_{s2}^*	5853 ± 15	5899	5846		5844	5878

Table 3.3: *Measured and predicted masses of ground state and excited B mesons.*

symmetry for the excited states. The leading corrections to the spectrum are inversely proportional to the heavy quark mass. They write the mass of a heavy-light meson in the form:

$$m(nL_J(j_q)) = m(1S) + E(nL(j_q)) + \frac{C(nL_J(j_q))}{m_Q} \ , \tag{3.24}$$

where $m(1S) = (3m(1S_1) + m(1S_0))/4$ is the mass of the ground state, $E(nL(j_q))$ is the excitation energy and $C(nL_J(j_q))$ are constants. The excitation energy $E(nL(j_q))$ has a weak dependence on the heavy quark mass. Their overall strategy is to use the observed properties of the K and D mesons to predict the properties of the orbitally excited B and B_s mesons. Given the fragmentary experimental knowledge of the reference systems, they referred to a non-relativistic (*Buchmüller type*) potential model to estimate the variation of the excitation spectrum as a function of heavy quark mass. The following mass predictions have been determined for orbitally excited B mesons:

$$\begin{aligned} m(B_1) &= 5755 \text{ MeV/c}^2 \\ m(B_2^*) &= 5767 \text{ MeV/c}^2 \\ m(B_{1s}) &= 5834 \text{ MeV/c}^2 \\ m(B_{2s}^*) &= 5846 \text{ MeV/c}^2 \ . \end{aligned} \tag{3.25}$$

Thus, the mass splitting between the 2^+ and the 1^+ state is predicted to be 12 MeV/c^2. Similar mass splittings are predicted by Falk and Mehen (3.22), although their absolute mass predictions are significantly different from the above.

Predictions on the basis of the *non-relativistic quark model* have been performed by Gupta and Johnson [GJ95] and quite recently by Isgur [Isg98]. In the paper of Isgur, a generic two-body effective potential for the $Q\bar{q}$ system of the form

$$V = V_c + V_{ss} + V_{so} \tag{3.26}$$

is used. The three terms are the central (c), the spin-spin (ss) and the spin-orbit (so) potentials. This ansatz is valid if both m_Q and m_q are large compared to Λ_{QCD}. Deviations from this ansatz are taken into account using a double heavy quark expansion to the order $1/m_Q^2$, $1/m_Q m_q$ and $1/m_q^2$. The standard assumption of the non-relativistic valence quark model is that the same Schrödinger equation, which *would* be valid for sufficiently heavy quarks Q and $\bar{q}$, may be extrapolated to constituent quark masses of the order of Λ_{QCD}. The spectra of $Q = u$ to $Q = b$ ($\pi - \rho$, $\bar{K} - \bar{K}^*$, $D - D^*$ and $\bar{B} - \bar{B}^*$) systems strongly suggest that the breaking of heavy quark symmetry is a smooth function of m_Q from $m_Q \to \infty$ to $m_Q = m_u$. Solving the Schrödinger equation for a known system results in the parameters of the potentials, defined above, at that particular mass scale. On the basis

of the observed smoothness of the parameter values with respect to m_Q, extrapolation of the fit parameters obtained at the known $\bar{K} - \bar{K}^*$ system to the $b\bar{d}$ system predicts the excited spectrum to be:

$$\begin{aligned} m(B_0^*) &= 5870 \text{ MeV/c}^2 \\ m(B_1) &= 5875 \text{ MeV/c}^2 \\ m(B_1) &= 5700 \text{ MeV/c}^2 \\ m(B_2^*) &= 5715 \text{ MeV/c}^2 \ . \end{aligned} \tag{3.27}$$

Checks on the validity of this approach are its correct prediction of the observed properties of the charm and bottom systems, as well as the comparison of its extrapolation to $Q = u$ with experimental results. The most striking result of Isgur is the *spin-orbit inversion*, namely that the $J_l^P = \frac{1}{2}^+$ states lie *above* rather than below the $J_l^P = \frac{3}{2}^+$ states. While contrary to atomic and $Q\bar{Q}$ spectra, this inversion has a simple interpretation in the quark model: Thomas precession in the spin-spin potential has exceeded the usual atomic-like spin-orbit force.

The mass spectrum of orbitally and radially excited states of B and D mesons is calculated taking into account an expansion in inverse powers of the heavy quark mass to first order. Non-relativistic quark model approaches also carry out an expansion in inverse powers of the light quark mass. Though, estimates of the light quark velocity in charm and bottom mesons show that the light quark is highly relativistic, $\beta \sim 0.75$, making a relativistic treatment of the light quark important. A *relativistic quark model* has been applied by Ebert, Galkin and Faustov [EGF98], [GEF02], by Di Pierro and Eichten [DPE01], by Godfrey and Kokolski [GK91] and by Zeng, Van Orden and Roberts [ZVOR95]. The predictions in the paper of Ebert, Galkin and Faustov are based on a quasi-potential approach in quantum field theory, successfully used for the calculation of heavy quarkonia spectra and electroweak decays of heavy mesons [FGM95, EFG97]. Quasi-potentials are mixtures of vector and scalar potentials, responsible for the observed asymptotic freedom and confining properties of QCD respectively. Their predictions for the mass differences between excited states of B and D mesons are:

$$\begin{aligned} \bar{m}_{B^{**}} - \bar{m}_{D^{**}} &= 3.29 \text{ GeV/c}^2 \\ \bar{m}_{B_s^{**}} - \bar{m}_{D_s^{**}} &= 3.30 \text{ GeV/c}^2 \\ \bar{m}_B - \bar{m}_D &= 3.34 \text{ GeV/c}^2 \\ \bar{m}_{B_s} - \bar{m}_{D_s} &= 3.33 \text{ GeV/c}^2 \ , \end{aligned} \tag{3.28}$$

where $\bar{m}_B$ and $\bar{m}_B$ are the appropriate spin averaged $P-$ and $S-$wave states of equation 3.20. The comparison of these numbers with the HQET prediction that they should be equal shows a high level of agreement.

The predicted masses for B states are:

$$\begin{aligned} m(B) &= 5285 \text{ MeV/c}^2 \\ m(B^*) &= 5324 \text{ MeV/c}^2 \\ m(B_0^*) &= 5738 \text{ MeV/c}^2 \\ m(B_1) &= 5757 \text{ MeV/c}^2 \\ m(B_1) &= 5719 \text{ MeV/c}^2 \\ m(B_2^*) &= 5733 \text{ MeV/c}^2 \end{aligned} \tag{3.29}$$

and for B_s states:

$$\begin{aligned} m(B_s) &= 5375 \text{ MeV/c}^2 \\ m(B_s^*) &= 5412 \text{ MeV/c}^2 \\ m(B_{s0}^*) &= 5841 \text{ MeV/c}^2 \\ m(B_{s1}) &= 5859 \text{ MeV/c}^2 \\ m(B_{s1}) &= 5831 \text{ MeV/c}^2 \\ m(B_{s2}^*) &= 5844 \text{ MeV/c}^2 \text{ .} \end{aligned} \tag{3.30}$$

Ebert, Galkin and Faustov find the same inversion of the broad and narrow states as Isgur, but the gap between the states is smaller.

An estimate of the excited B spectrum without resort to modelling is given by a *lattice calculation* using the non-relativistic QCD (NRQCD) formulation for heavy quarks and the tadpole-improved clover action for light quarks by A. Ali Khan and co-workers [AK+00]. This approach has allowed the authors to have better control over discretization errors in both the heavy and the light quark sectors. The NRQCD approach is justified because the typical velocity of the heavy quark is small: $\beta \sim \Lambda_{QCD}/m_Q \sim 0.05$. The standard value of lattice scale used was $1/a = 1.927$ GeV. The predicted B meson masses are:

$$\begin{aligned} m(B) &= 5296 \pm 4^{+3}_{-2} \text{ MeV/c}^2 \\ m(B^*) &= 5319 \pm 2^{+3}_{-2} \text{ MeV/c}^2 \\ m(B_0^*) &= 5670 \pm 37^{+16}_{-24} \text{ MeV/c}^2 \\ m(B_2^*) &= 5822 \pm 45^{+27}_{-35} \text{ MeV/c}^2 \end{aligned} \tag{3.31}$$

and for B_s states:

$$\begin{aligned} m(B_s) &= 5385 \pm 15\ ^{+7}_{-6}\ ^{+20}_{-0} \text{MeV/c}^2 \\ m(B_s^*) &= 5412 \pm 2\ ^{+2}_{-4}\ ^{+20}_{-0} \text{MeV/c}^2 \\ m(B_{s0}^*) &= 5742 \pm 27\ ^{+14}_{-20}\ ^{+15}_{-0} \text{MeV/c}^2 \\ m(B_{s2}^*) &= 5878 \pm 26\ ^{+23}_{-33}\ ^{+11}_{-0} \text{MeV/c}^2 \text{ ,} \end{aligned} \tag{3.32}$$

where the first error is statistical, the second comes from varying a^{-1} between 1.8 and 2.0 GeV, and for the strange mesons, the third error comes from the uncertainty in the strange quark mass.

3.7.2 The Prediction of Decay Widths

On the one hand, it is somewhat more delicate to make predictions for the widths of the excited B mesons than for their masses, since the widths depend on the available phase space, and hence on the values of the heavy meson masses. On the other hand, the widths of the broad and narrow states can be estimated by simple quantum mechanical considerations. The broad states are expected to have a width of a few hundred MeV and the narrow states a width of a few MeV. An uncertainty on the absolute mass of the states of the order of ~ 20 MeV would change the width of the states by a few MeV. The broad states would remain broad and would still be a challenge to analyse. The narrow states could get broader or narrower by a few percent. Experimentally, there is essentially no sensitivity to the change in the width of the states.

As a result, only the prediction of one paper for the width of the excited B states is presented here. The derivation of the strong decay widths of excited heavy hadrons may be found in [MW00]. Di Pierro and Eichten predict the decay widths of the $B^{**}_{u,d}$ and the B^{**}_s mesons, using a relativistic quark model, to be [DPE01]:

Decay Channel		Width Γ	p_x	Partial Wave
$\Gamma(B_0^* \to B\pi)$	=	125	388	S
$\Gamma(B_1 \to B^*\pi)$	=	121	381	S
$\Gamma(B_1 \to B^*\pi)$	=	8.8	338	D
$\Gamma(B_2^* \to B\pi)$	=	7.1	396	D
$\Gamma(B_2^* \to B^*\pi)$	=	6.4	352	D
$\Gamma(B_{s0}^* \to BK)$	=	107	170	S
$\Gamma(B_{s1} \to B^*K)$	=	96	153	S
$\Gamma(B_{s1} \to B^*K)$	=	0.01600		D
$\Gamma(B_{s2}^* \to BK)$	=	0.00064		D
$\Gamma(B_{s2}^* \to B^*K)$	=	0.00962		D

Hadronic transitions for B and B_s systems. Values for p_x and Γ are given in MeV. Since the narrow B_s^{**} states are very near the kinematical threshold, the decay widths have their phase space dependence divided out. For these states the numbers correspond to $\Gamma \times (100/p_x)^{2l_x+1}$.

Chapter 4

B Spectroscopy

The search for and analysis of orbitally excited B mesons is the main goal of this analysis. The narrow B_u^{**} and B_d^{**} states have been experimentally established, whereas for the corresponding broad states no results have been obtained. The evidence for the B_s^{**} is less compelling and, to date, OPAL is the only collaboration reporting a significant signal [A$^+$95c].

After giving a general description of the experimental method of B spectroscopy, this chapter summarizes the experimental results in the field of excited B mesons, putting the results of this analysis into perspective. The analogous results in the charm sector are a motivation for upcoming analyses and also give a hint towards the validity of theoretical predictions for excited heavy mesons. The present status of excited D meson spectroscopy is given in the last section of this chapter.

4.1 The Experimental Method

Excited states under examination have due to their connstituent quarks and excitation energy a certain mass. The reconstruction of this mass via their decay products is the aim of B spectroscopy. Due to the finite, although very short, lifetime of the excited states the masses are not exactly defined but form resonances. The functional form can be described by the Breit-Wigner function, which is the fourier transform of the exponentially falling lifetime distribution of a given quantum mechanical state. The experimental result of a measurement concerning a decaying state is distorted in two ways. First, by the finite resolution of the detector smearing out physical structures, thus resulting in a rather Gaussian form of a resonance, provided the width of the resolution is comparable or larger than the natural width of the resonance. Second, background events contributing to the mass distribution have no correlation to the decaying particle and consequently do not form a resonance. In summary, the task of a spectroscopy analysis is to enrich and examine a resonance peak in the mass distribution which

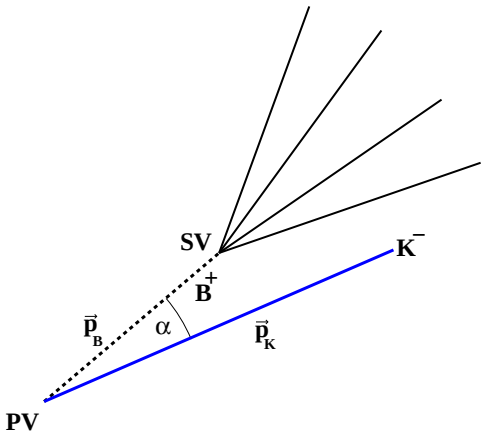

Figure 4.1: *A schematic view of a B_s^{**} decay into a B^+ and a K^-.*

generally is located on a certain background form.

For the generic 2-body decay chain,

$$B^{**} \to B^{(*)} T_{prim} \tag{4.1}$$

where B^{**} represents the orbitally excited state, $B^{(*)}$ the b-hadron decay product and T_{prim} the associated primary track, the invariant mass of the $B^{(*)}T_{prim}$ pair is defined by:

$$M \equiv m(B^{(*)}T_{prim}) = \sqrt{m_B^2 + m_T^2 + 2(E_B E_T - \vec{p}_B \vec{p}_T \cos\alpha)} \tag{4.2}$$

where m_B, m_T, E_B, E_T, $\vec{p}_B$ and $\vec{p}_T$ have been used for the masses, energies and momenta of the B meson and the associated primary track. α denotes the opening angle between the two decay products in the lab frame and has a resolution that is mainly governed by the direction resolution of the b-hadron (see figure 4.1).

The most natural quantity with which to study B spectroscopy is not the reconstructed invariant mass of the state directly, but rather the Q-value. The Q-value is defined as:

$$\begin{aligned} Q &= m(B^{**}) - m(B^{(*)}) - m(T_{prim}) \\ &= m(B^{(*)}T_{prim}) - m(B^{(*)}) - m(T_{prim}) \end{aligned} \tag{4.3}$$

and is a measure for the kinetic energy of the decay products. The lowest value for the Q-value is thus zero. In this case the decay particles have no momentum in the rest frame of the mother particle.

The resolution of Q assuming independent variables is given by usual error propagation by:

$$\sigma_Q = \sqrt{\left(\frac{\partial Q}{\partial p_B}\right)^2 \sigma_{p_B}^2 + \left(\frac{\partial Q}{\partial p_T}\right)^2 \sigma_{p_T}^2 + \left(\frac{\partial Q}{\partial \alpha}\right)^2 \sigma_\alpha^2} \,. \tag{4.4}$$

This has proven to be incorrect since the variables are correlated and the distributions are non-Gaussian. Estimates of individual errors attached to specific combinations of a B and a primary track candidate is provided by a neural network approach. This technique takes into account all variable correlations and is sensitive to the non-linear nature of the distributions. The details of the neural network ansatz are further explained in chapter 7.

Primary track candidates T_{prim} are expected to have typical momenta of a few GeV. In this range, single track resolution is of the order of a percent with an angular resolution less than 1 mrad. σ_{p_T} is thus negligible compared to σ_{p_B} and σ_α. The effect of measurement errors regarding B energy and B momentum direction can be seen in figure 4.2. The plots show that the measurement of the B direction and thus also of α are dominating the Q-value resolution.

The case where the b-hadron was a B^* meson which decayed into a photon, which was not reconstructed, has a negligibly small effect on the Q-value resolution. However, the decay of a B^{**} of a given mass to B^*T_{prim} gives rise to a Q-value which is shifted downwards by the $B^* - B$ mass difference of approximately 45 MeV/c^2 compared to the Q-value of a BT_{prim} decay. Therefore, in order to extract the B^{**} mass from the measured Q-value, an assumption about the B^*/B ratio in the signal has to be made.

The evaluation of the Q-value in equation 4.3 requires the measurement of the track T_{prim} and the reconstruction of the b-hadron decay product 4-momentum. There are two main approaches to achieve this task: the exclusive and the inclusive approach. The exclusive approach has the advantage of high measurement precision and low background contamination but the disadvantage of low statistics, due to the huge number of B decay channels. Only a very little fraction of the possible decay products is suitable for the analysis limiting the available statistics. Inclusive methods have no problems with low statistics, the drawback is the rather un-precise measurement of the B 4-momentum. Analyses based on both approaches have acquired results on the spectroscopy of B mesons, and will both be described in the next section.

4.2 Experimental Results on Excited b-Hadrons

4.2.1 Inclusive Studies

Evidence for the existence of $B^{**}_{u,d}$ states first emerged in analyses at LEP in which a charged pion produced at the primary event vertex was combined with an inclusively reconstructed B-meson [A^+95b, B^+96, A^+95c]. Common to the experimental methods was:

- A b enhanced sample of hadronic events was selected by applying a lifetime based event tag.

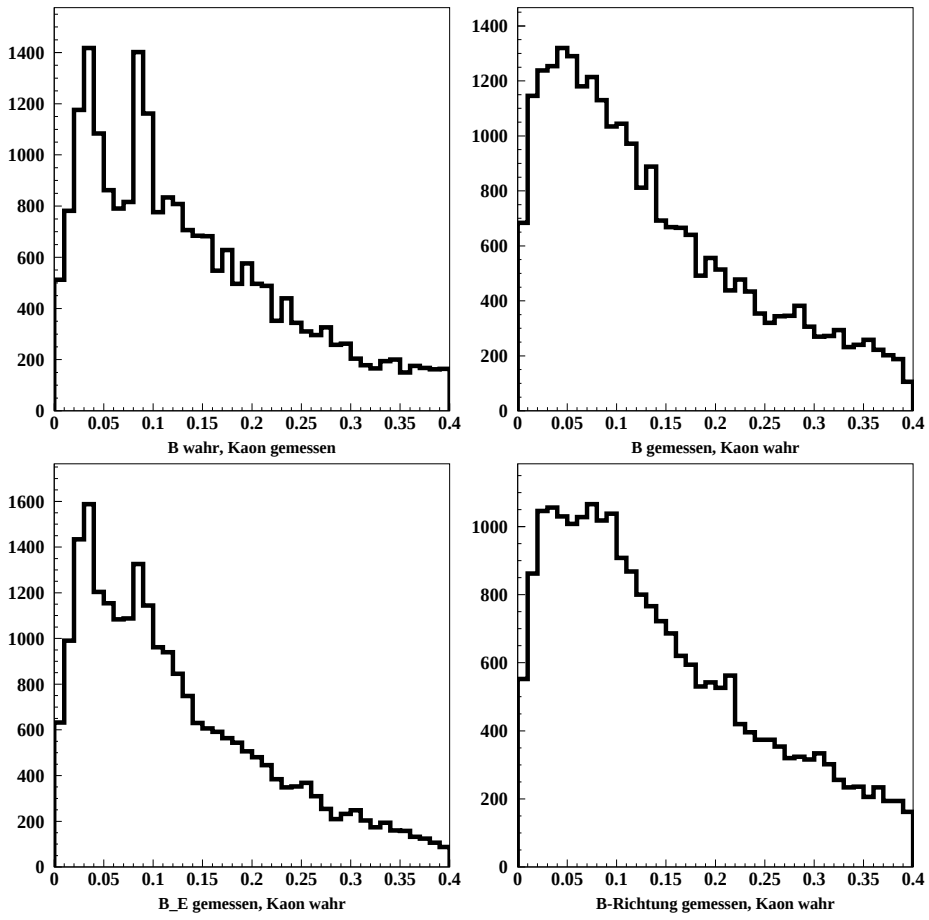

Figure 4.2: *Effects of measurement errors on the Q-value resolution. The plots show the effect of measuring the kaon momentum (upper left), the B momentum. Clearly, the effect of the uncertainty in the B measurement dominates. Lower left is the effect of the uncertainty in the B energy and right in the B momentum direction.*

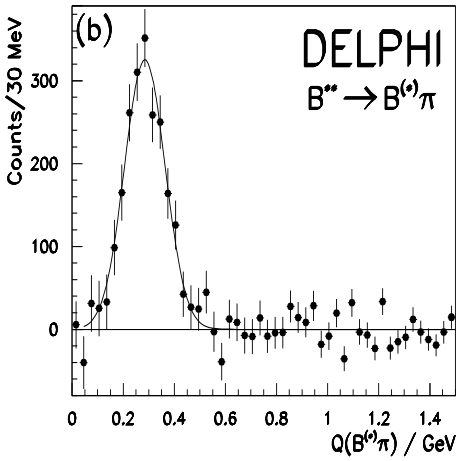

Figure 4.3: *Measured Q-value distribution of the DELPHI collaboration.*

- The 4-momentum of the B meson was estimated using the kinematical information from the secondary vertex and from the tracks with large rapidity with respect to the b-jet axis.
- The primary track from the B^{**} decay was identified as the primary vertex charged track (oppositely charged to the decay b-hadron) with the largest rapidity along the B direction.

The signal has been observed as a peak in the Q-value distribution, with a typical mass resolution of 50 MeV. No attempt has been made to disentangle the structure of the B^{**} resonance. After background subtraction, the fits to the signal gave the masses and production rates:

$$\begin{array}{rrcl} \text{OPAL} & m_{B^{**}} & = & 5681 \pm 11 \text{ MeV} \\ & f_{B^{**}} & = & 0.270 \pm 0.056 \\ \text{DELPHI} & m_{B^{**}} & = & 5734 \pm 5 \pm 17 \text{ MeV} \\ & f_{B^{**}} & = & 0.320 \pm 0.018 \pm 0.06 \\ \text{ALEPH} & m_{B^{**}} & = & 5703 \pm 4 \pm 10 \text{ MeV} \\ & f_{B^{**}} & = & 0.279 \pm 0.016 \pm 0.059 \ . \end{array} \tag{4.5}$$

In the analysis published by the L3 Collaboration [A$^+$99], an attempt has been made to resolve the structure of the $B^{(*)}\pi$ mass spectrum. According to spin-parity rules, one expects resonances of five possible B^{**} decay modes: $B_2^* \to B\pi$, $B_2^* \to B^*\pi$, $B_1 \to B^*\pi$, $B_1 \to B^*\pi$ and $B_0^* \to B\pi$. The signal has been fitted using five Voigt functions to account for effects due to detector resolution. The constraints imposed rely on predictions common to existing HQET models which are based on the phenomenology of orbitally excited D mesons [A$^+$86, A$^+$89a, A$^+$89b, A$^+$96b]:

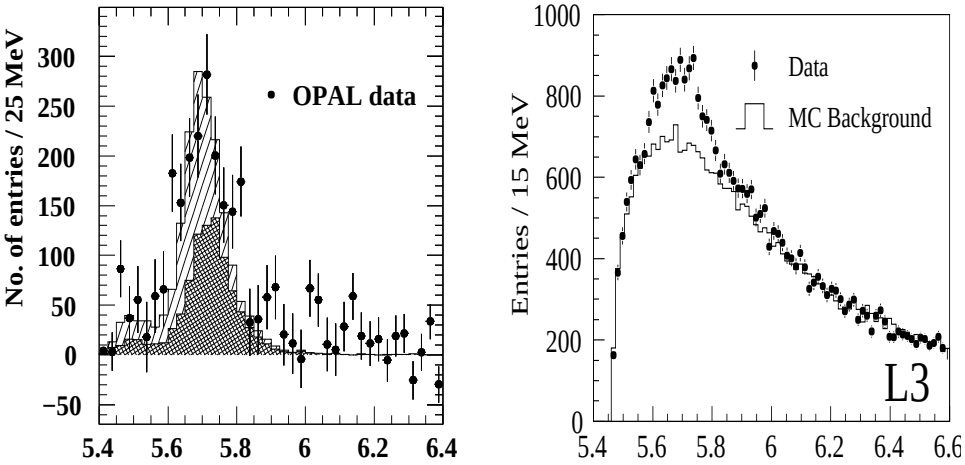

Figure 4.4: *Measured B^{**} Q-value distribution of the OPAL and L3 collaborations.*

- By extrapolating from measurements in the D sector, one can expect that the widths of the $j_{light} = 3/2$ states are approximately equal and in the range of $20 - 25$ MeV. For the case of the $J_{light} = 1/2$ states, there were in 1998 no corresponding measurements in the D system and so the prediction is only based on the theoretical expectation that the widths are approximately equal albeit rather broader than for the $J_{light} = 3/2$ states.

- According to HQET predictions [EHQ93, EGF98], the mass difference between the two $J_{light} = 1/2$ and the two $J_{light} = 3/2$ states should be approximately equal and in the range of $5 - 20$ MeV.

- The absolute size of the masses of the $j_q = 3/2$ states are expected to lie $\sim$100 MeV higher than the masses of the $j_q = 1/2$ states, in consensus to predictions of [GR94, GJ95], and in disagreement with [Isg98, EGF98] predicting a spin-orbit inversion.

To summarize, the L3 analysis imposes the constraints that $\Gamma_{B_1} = \Gamma_{B_0^*}$ and $\Gamma_{B_1} = \Gamma_{B_2^*}$ plus $M_{B_1} - M_{B_0^*} = 12$ MeV and $M_{B_2^*} - M_{B_1} = 12$ MeV. Thus, there are 4 fitted parameters: the widths and masses of the states B_1 and B_2^*. The results obtained are:

$$\begin{aligned} m_{B_1} &= 5670 \pm 10(\text{stat.}) \pm 13(\text{syst.}) \text{ MeV} \\ \Gamma_{B_1} &= 70 \pm 21(\text{stat.}) \pm 25(\text{syst.}) \text{ MeV} \\ m_{B_2^*} &= 5768 \pm 5(\text{stat.}) \pm 6(\text{syst.}) \text{ MeV} \\ \Gamma_{B_2^*} &= 24 \pm 19(\text{stat.}) \pm 24(\text{syst.}) \text{ MeV} \end{aligned}$$

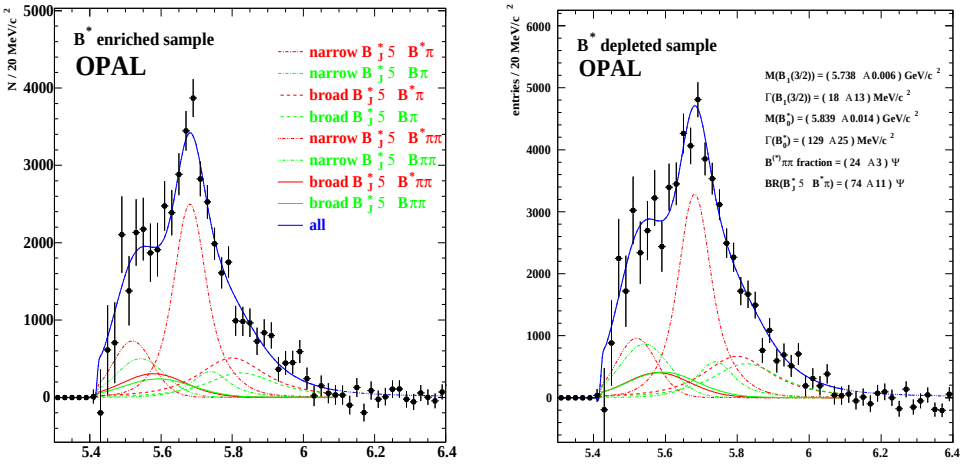

Figure 4.5: *Measured B_d^{**} Q-value distribution of the OPAL collaboration.*

with an additional excess of signal events near 5.9 – 6.0 GeV, which is interpreted as resulting from the decay $B' \to B^{(*)}\pi$. The B' mass, Gaussian width and branching fraction are found to be:

$$\begin{aligned} m_{B'} &= 5937 \pm 21 \pm 4 \text{ MeV} \\ \sigma_{B'} &= 50 \pm 22 \pm 5 \text{ MeV} \\ f_{B'} &= 0.034 \pm 0.011 \pm 0.008 \ . \end{aligned}$$

These results disfavour recent theoretical models proposing spin-orbit inversion, but agree well with earlier models.

In a recent paper [A+02], the OPAL Collaboration presented the first measurement of the branching ratio of orbitally excited B mesons decaying into $B^*\pi X$. In order to disentangle $B^{**} \to B^*\pi$ from $B^{**} \to B\pi$ decays, OPAL uses the information of the photon in the decay $B^* \to B\gamma$. For each B meson candidate, a probability weight $\mathcal{W}(B^*)$ is assigned to come from a B^* decay. The probability weight is used to obtain complementary subsamples enriched or depleted in their B^* content. From the multiplicities of the subsamples, the model-independent result for the branching ratio of orbitally excited B mesons decaying to B^* is found to be:

$$\text{BR}(B^{**} \to B^*\pi(X)) = 0.85 \ ^{+0.26}_{-0.27} \pm 0.12 \qquad (4.6)$$

in agreement with theoretical predictions [GR94]. In the framework of HQET, a simultaneous fit to the $B\pi$ background-subtracted and efficiency-corrected mass spectra of the B^* enriched and depleted subsamples yields the mass and width of the B^{**} mesons as well as the branching ratios into

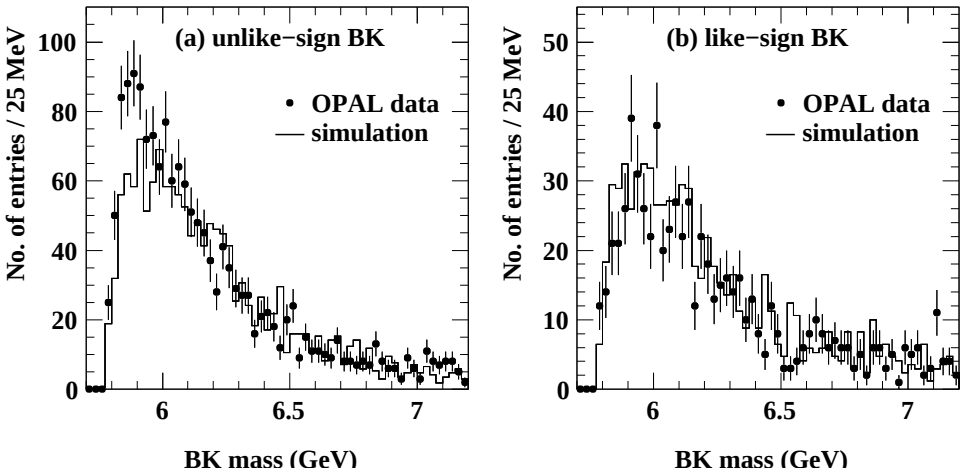

Figure 4.6: *Measured Q-value distribution of the OPAL collaboration.*

$B^*\pi$ and $B^{(*)}\pi\pi$:

$$\begin{aligned} m_{B_0^*} &= 5839\ ^{+0.013}_{-0.014}\ ^{+0.034}_{-0.042}\ \mathrm{MeV}/c^2 \\ \Gamma_{B_0^*} &= 129\ ^{+27}_{-23} \pm 63\ \mathrm{MeV}/c^2 \\ m_{B_1} &= 5738\ ^{+0.005}_{-0.006} \pm 0.007\ \mathrm{MeV}/c^2 \\ \Gamma_{B_1} &= 18\ ^{+15}_{-13}\ ^{+29}_{-23}\ \mathrm{MeV}/c^2 \\ \mathrm{BR}(B^{**} \to B^*\pi) &= 0.74\ ^{+0.12}_{-0.10}\ ^{+0.21}_{-0.15}\ . \end{aligned} \tag{4.7}$$

The properties of the doublet partners of the B_0^* and B_1 states are constrained in the fit. The constraints are similar to the ones applied in the L3 analysis: $\Gamma_{B_1} = 1.25 \cdot \Gamma_{B_0^*}$, $\Gamma_{B_2^*} = \Gamma_{B_1}$, $M_{B_1} - M_{B_0^*} = 20$ MeV and $M_{B_2^*} - M_{B_1} = 12$ MeV. The production ratio of the states was set to $f(B_0^*) : f(B_1)_{1/2} : f(B_1)_{3/2} : f(B_2^*) = 2 : 2 : 3 : 3$ corresponding to the average production rate according to state counting, total spin counting and light quark spin counting. To justify this constraint, a separate fit to the total $B\pi$ spectrum has been performed with f_{narrow} as an additional free parameter, giving the result:

$$f_{\mathrm{narrow}} = 0.76\ ^{+0.11}_{-0.24}\ \mathrm{stat} \tag{4.8}$$

which is in agreement with the fixed value of 0.6 used in the original fit.

For the case of the B_s^{**}, if the meson mass is above the $B^{(*)}K$ threshold, this will be the dominant decay mode since the $B_s\pi$ channel is forbidden by isospin invariance. The predictions from HQET presented in Table 3.2 for $B_{u,d}^{**}$ now carry over with the replacement of the primary decay pion by a kaon. Experimentally, evidence for the B_s^{**} is less compelling than for the

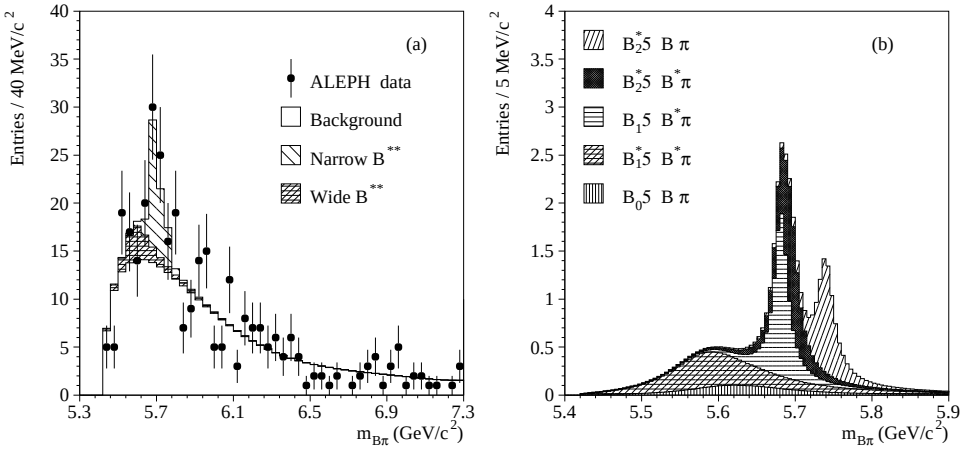

Figure 4.7: *(a) The $B\pi$ mass spectrum from ALEPH data (points with error bars) and fit (histogram). The fit includes the expected background plus contributions from the narrow and wide B^{**} states. (b) An expanded view of the signal region, showing the contributions to the $B\pi$ mass spectrum from the five B^{**} decay channels.*

$B^{**}_{u,d}$ and, to date, OPAL [A$^+$95c] is the only collaboration reporting a significant signal. They see an excess over background of 149 ± 31 B^+K^- pairs (with invariant masses in the range 5.80-6.00 GeV), which translates into,

$$\frac{BR(Z^0 \to \bar{b} \to B_s^{**0} \to B^{(*)+}K^-)}{BR(Z^0 \to \bar{b} \to B^+)} = 0.026 \pm 0.008 \ .$$

4.2.2 Exclusive Studies

Only the ALEPH Collaboration of the LEP experiments has performed an exclusive study of the B^{**} resonance [B$^+$98], based on the samples of B mesons reconstructed in the decay modes: $B \to D^{(*)}\pi^{\pm}(\rho^{\pm}, a_1^{\pm})$ and $B \to J/\Psi(\Psi')K^{\pm}(K^0)$. The selection gives 238 charged and 166 neutral B meson candidates. In order to reconstruct the B^{**} states, the B mesons are combined with a charged pion coming from the primary vertex. The achieved mass resolution is better than 5 MeV. An excess of 45 ± 13 events in the right-sign combinations is observed above the expected background estimated from wrong-sign combinations. Figure 4.7 shows the fit to the right-sign mass spectrum where the signal shape consists of five Breit-Wigner functions. Because of the limited statistics, however, only the mass of the B_2^* state and the relative production rate of the B^{**} system are left free in the fit, the fine and hyperfine splittings being fixed to their theoretical values predicted by heavy quark symmetry [EHQ93]. The results of the fit

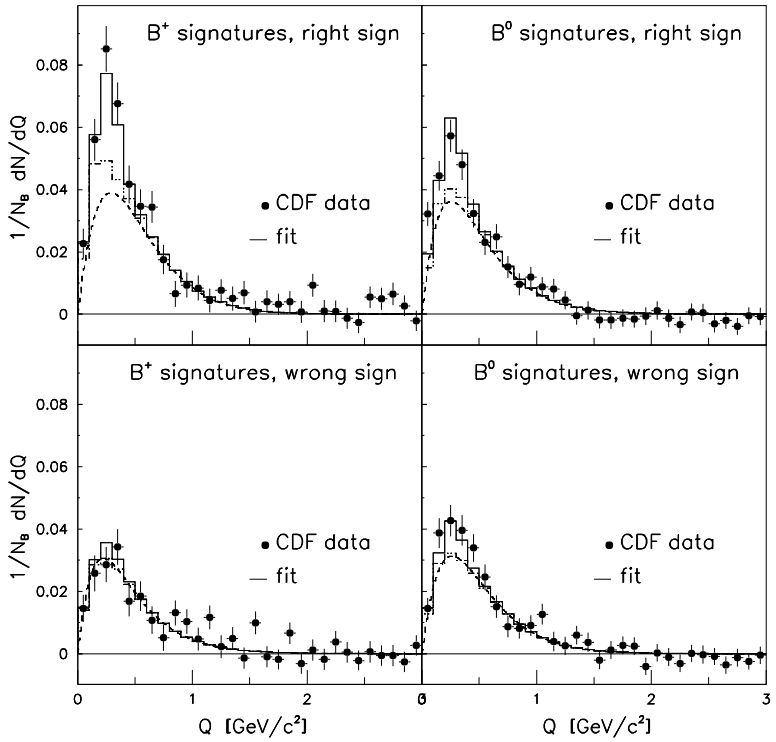

Figure 4.8: *The sideband subtracted $B\pi^{\pm}$ Q distribution of the data (points) compared to the fit results as obtained by the CDF collaboration. The dashed curves are the fitted hadronization component, the dotted histograms include all backgrounds, and the solid histograms are the totals including the B^{**} signal.*

are:

$$\begin{aligned} m_{B_2^*} &= 5739\ ^{+8}_{-11}\ ^{+6}_{-4}\ \mathrm{MeV}/c^2 \\ f_{B^{**}} &= 0.31 \pm 0.09\ ^{+0.06}_{-0.05} \end{aligned}$$

where the first error is statistical and the second is due to systematics.

More recently, the CDF Collaboration has presented the first results on $B^{**}_{u,d}$ production from a hadron collider [A$^+$01c] by combining a charged primary pion to a partially reconstructed B using semileptonic decays to charm $B \to l\bar{\nu}\bar{D}^{(*)}$. This decay channel is a natural choice since the lepton provides an easy trigger for the event. The $D^{(*)}$ meson is fully reconstructed in several decay modes. Both the production and decay vertices of the B are reconstructed and used to estimate its flight direction, while its momentum is taken from the decay products, after rescaling by 15% for the neutrino. This partial reconstruction also allows the identification of the charge and flavour of the B meson. The achieved Q-value resolution is 50 MeV.

The Q-value distributions are separated into a right-sign and wrong-sign subsample corresponding to the π charge to B flavour correlation. After the selection, an excess of signal is observed in the right-sign subsample. In order to extract the signal, several background sources have been considered. The Q distribution after the subtraction of uncorrelated backgrounds is shown in figure 4.8. The correlated background, i.e. pions produced next to the B in the fragmentation, is taken into account using Monte Carlo simulation. The parametrized background shape of this background source is fitted to the data together with the signal. The fit to the signal yields:

$$\begin{aligned} m_{B_1} &= 5710 \pm 20 \text{ MeV}/c^2 \\ f_{B^{**}} &= 0.28 \pm 0.06 \ \pm 0.03 \end{aligned}$$

where the mass splittings between the four states is assumed to be as predicted in [Isg98].

4.2.3 Excited Baryon Studies

No detailed numbers are known for the case of the primary hadron being a baryon, but flavour and spin counting rules indicate that a large fraction of primary b-baryons will be excited states e.g. Σ_b, Σ_b^* which will decay by either the electromagnetic or strong interaction into the lightest b-baryon state, the Λ_b. In addition to the Λ_b, probably the only other weakly decaying b-baryons are the Ξ_b and the Ω_b. Experimentally, there are currently no confirmed signals reported for the $\Sigma_b^{(*)}$ although there is some evidence for Ξ_b production [BLMÅ02].

4.3 Experimental Results in the Charm Sector

Spectroscopy in the charm sector is identical to the technique used for B spectroscopy. HQET is an effective theory that holds in the limit of infinite heavy quark mass. Since the mass of the c quark is lower than that of the b quark, the mass dependent corrections on the predictions of HQET are different. The predicted mass spectrum of the D system can be summarized as in figure 4.9.

The D and D^* mesons, in which the light quark has orbital angular momentum $L = 0$, are well established and their properties and decays have been extensively studied [H$^+$02].

Excited D mesons with orbital angular momentum $L = 1$ are collectively called D^{**}. The observed $D_1(2420)$ and $D_2^*(2460)$ are interpreted as the spin doublet with $j_q = \frac{3}{2}$ [H$^+$02]. The first observation of the wide D_1^0 has been recently reported by CLEO II [A$^+$00b]. The mass of the D_1^0 meson has been measured to be $m_{D_1^0} = 2461^{+41}_{-34} \pm 10 \pm 32$ MeV and its width to be $\Gamma_{D_1^0} = 290^{+101}_{-79} \pm 26 \pm 36$ MeV. The result is in agreement with HQS theory,

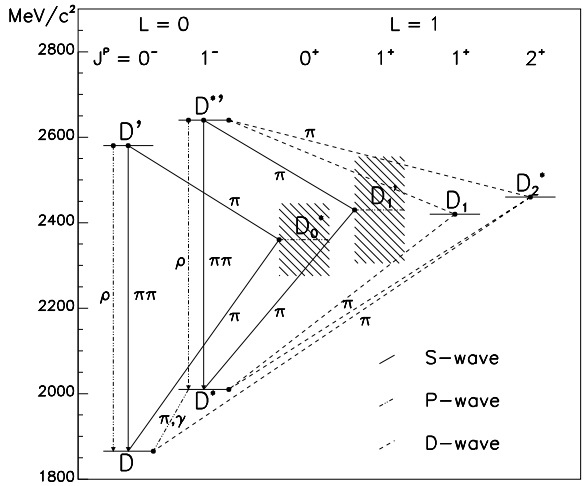

Figure 4.9: *The predicted D meson mass spectrum.*

but does not yet allow discrimination between different models (spin-orbit inversion or not). There has been no measurement made on the D_0^* state yet.

A still controversial issue is the observation of the radial excitation $D^{*\prime}$ claimed by DELPHI [A$^+$98], because similar analyses from OPAL [A$^+$01a] and CLEO [Rod98] found no evidence.

In the strange sector of charmed mesons, in addition to the $L = 0$ states, the $D_s^{\pm}$ and $D_s^{*\pm}$, two narrow resonances, the $D_{s1}^{\pm}(2536)$ and the $D_{sJ}^{\pm}(2573)$, compatible with the $L = 1$, $j_q = \frac{3}{2}$ doublet have been found [H$^+$02], but no new data are available.

Chapter 5

The DELPHI Detector

DELPHI, DEtector with Lepton, Photon and Hadron Identification, is a detector for e^+e^- physics, with special emphasis on particle identification, three-dimensional tracking information, high granularity and precise vertex information. It was installed at the Large Electron and Positron collider (LEP) at CERN, the European Laboratory for Particle Physics close to Geneva, where it has operated from 1989 to 2000.

LEP had a circumference of 26.7 km and was situated in the LEP tunnel, about 100 m below ground. Four of the eight interaction regions were equipped with detectors of the ALEPH, L3, OPAL and DELPHI collaborations.

In the first phase of LEP, lasting from 1989 to 1995, electrons and positrons collided at a centre-of-mass energy of about 91 GeV, so that in these collisions a Z^0 particle would be produced. Electrons and positrons enter LEP through a multistage accelerator system with an energy of 20 GeV. In LEP the particles were accelerated by 128 copper radio-frequency field cavities, which have been gradually replaced by superconducting cavities. This and the fast development of accelerating techniques have enabled a second phase of LEP, called LEP2, with a gradual increase of the centre-of-mass energy, first above the W^+W^- pair production threshold during 1996 and at last to 209 GeV in the year 2000.

LEP has been an excellent place to measure the characteristics of the Z^0 boson and its decays due to the vast increase in production rate at the Z^0 resonance. The major achievements of LEP have been to determine the number of light neutrino species to be three and to verify the Standard Model with high precision. The major achievements of LEP2 have been to measure the characteristics of the W boson. Beside this, LEP2 has played a major role in the search for new particles, in particular the Higgs boson and supersymmetric particles.

In the standard DELPHI coordinate system, the z axis is along the electron direction, the x axis points toward the centre of LEP, and the y axis

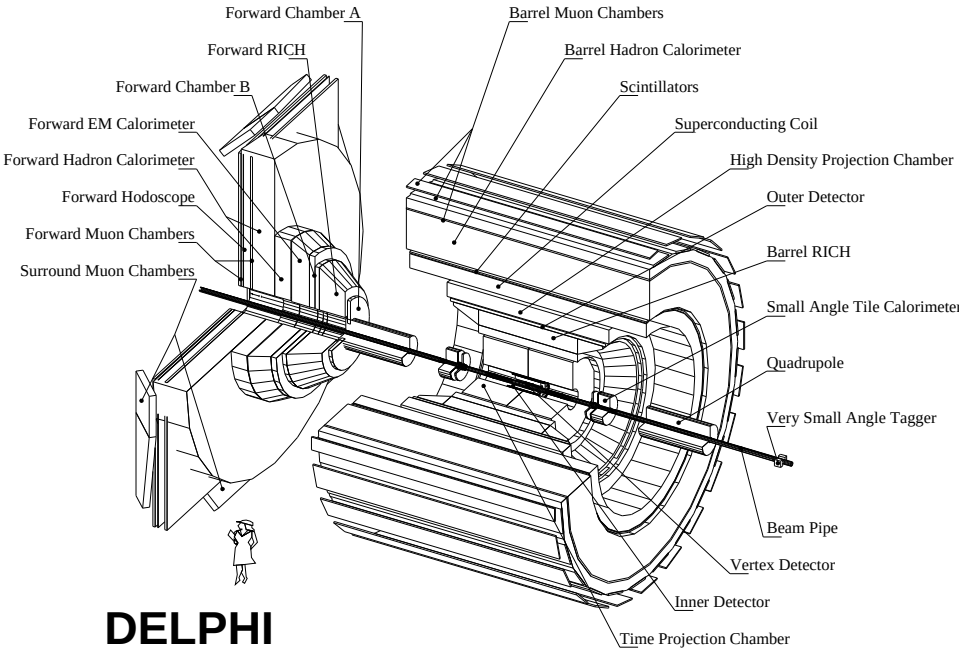

Figure 5.1: *Schematic view of the DELPHI detector.*

points upwards. The polar angle to the z axis is called θ and the azimuthal angle around the z axis is called ϕ. The radial coordinate is $R = \sqrt{x^2 + y^2}$.

The DELPHI detector has been described in detail elsewhere [DEL95]. Only the components relevant for the analysis described in this thesis will be discussed. The detector consists of a cylindrical section covering the 'barrel' region and two end-caps covering the 'forward' regions. Figure 5.1 schematically shows the layout of the barrel and of one end-cap.

5.1 The Tracking System

Charged particle tracking was performed by the Vertex Detector (VD), the Inner Detector (ID), the Time Projection Chamber (TPC) and the Outer Detector (OD). In the following, the performance achieved by each of these is briefly reviewed.

Spectroscopy of b-Hadrons requires a high precision tracking since the quality of the reconstruction of the primary and secondary B states is reliant on the precision of the single track measurement. This is all the more important as, for the high boost at LEP energies the angles of the tracks with respect to the jet axis get very small in the laboratory frame.

Tracking devices like the Forward Chamber A and B are not reviewed in this section since they have not been used for this analysis.

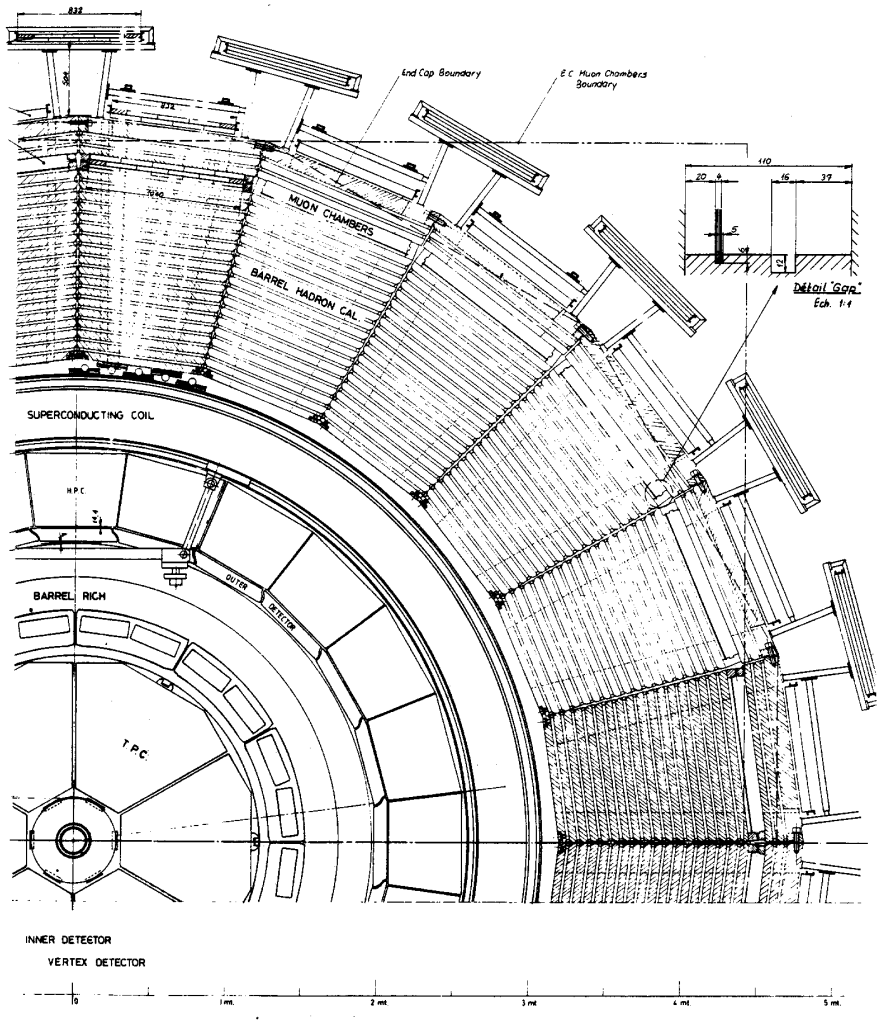

Figure 5.2: *Schematic view of the* DELPHI *detector (barrel part) along the beam pipe.*

5.1.1 The Solenoid

A large part of the barrel detector is embedded in the magnetic field (1.2T) of the large superconducting coil of 7.2 m length and an inner diameter of 5.2 m. The conductor, made of copper-packed Nb-Ti, is operated at a temperature of $T = 4.5$ K with a total current of 5000 A. The magnetic field allows an accurate momentum determination of charged particles from the curvature of their trajectory. To ensure homogeneity of the field, compensating coils are mounted on both end flanges of the main coil. The magnetic field also plays an important role for the sub-detectors the design of which is based on charge drift in the z-direction. The drift parallel to the magnetic field strongly reduces transverse diffusion. This is the case for the Time Projection Chamber, the barrel Ring Imaging Cherenkov counter as well as the High density Projection Chamber. The radial component of the magnetic field is of the order of $4 \cdot 10^{-4}$.

5.1.2 The Vertex Detector

The DELPHI Vertex Detector, figure 5.3, is one of the most important detector components for B physics. The VD consists of three coaxial cylindrical layers of silicon micro-strip detectors which surround the interaction point at average radii of 6.3, 9.0 and 10.9 cm. It allowed high precision track measurements necessary for the reconstruction of secondary vertices.

Each layer consists of 24 modules covering a length of 23 cm (27 cm for the closer layer). In 1994, two layers of the VD were equipped with double-sided micro-strip detectors which provide measurements in $R\phi$ and z direction.

The intrinsic resolution of the VD for single hit tracks in the $R\phi$ direction is 7.6μm. The single hit precision of the z coordinate is a function of the incident angle of the track reaching a value of 9μm for tracks perpendicular to the module.

The short lever arm to the interaction point results in an excellent impact parameter[1] resolution in $R\phi$ and z. For tracks with hits in all three $R\phi$ VD layers the impact parameter resolution was

$$\sigma^2_{R\phi} = ([61/(P \sin^{3/2}\theta)]^2 + 20^2)\ \mu\mathrm{m}^2$$

and for tracks with hits in both Rz layers and with polar angle

$$\theta \approx 90^\circ, \sigma^2_{Rz} = ([67/(P \sin^{5/2}\theta)]^2 + 33^2)\ \mu\mathrm{m}^2$$

To achieve the extremely high precision a careful alignment of each layer is necessary. Measurements on high momentum particle trajectories (e.g.

[1] The impact parameter is defined as the distance of closest approach of a track to the primary vertex.

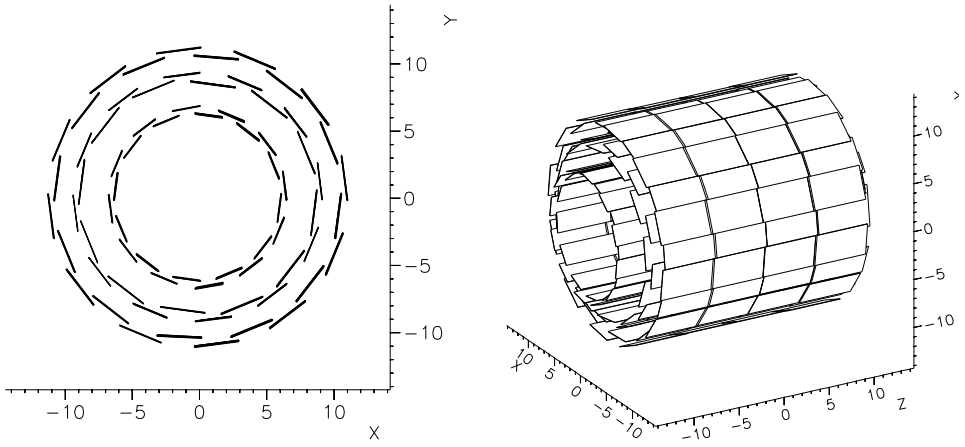

Figure 5.3: *Schematic layout of the micro vertex detector (x, y, z in cm). Left: projection along the beam pipe. Right: Perspective view.*

$Z^0 \to \mu^+\mu^-$) as well as the overlaps between the neighbouring wafers in each layer are very useful in this respect. Particle identification with the VD can be performed using the dE/dx measurement for charged tracks. This is described in detail in section 7.2.1.

5.1.3 The Inner Detector

The Inner Detector is a tracking and triggering detector. It consists of two parts: the Jet Chamber and the Trigger Layers.

The Jet Chamber is a drift chamber, subdivided into 24 sectors of 15° in ϕ. Each sector consists of 24 sense wires, measuring the drift time. The drift gas used is CO_2 with a small admixture of isobutane (C_4H_{10}, 4.7%) and isopropanol (C_3H_7, 0.7%). The drift voltages are chosen such that they increase linearly in r, i.e. the longer the drift distance, the higher the field. Therefore, the ionisation electrons from tracks originating from the interaction point all arrive approximately at the same time at one of the 24 wires. This feature is exploited for the fast trigger, which typically needs 5 μs to decide if a track originated from the interaction point or not.

The Trigger Layers consist of 5 cylindrical layers of Multi-Wire-Proportional-Chambers (MWPC). They are filled with an admixture of argon (70%) and CO_2 (30%). Mounted along the beam-pipe, each layer consists of 192 sense wires, 192 field-shaping wires and 192 circular cathode strips (5 mm pitch). The sense wires solve the left-right ambiguity of the Jet Chambers whereas the cathode strips provide z-information. After corrections for effects such as temperature and pressure variations, the resolution for muons from $Z^0 \to \mu^+\mu^-$ is 50 μm in $r\phi$ and 1.5 mrad in ϕ [DEL95]. The 2-track separation was of the order of 1 mm. The z precision from a single

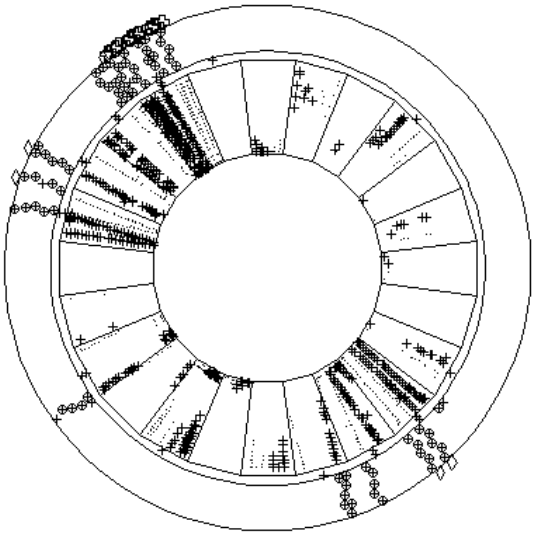

Figure 5.4: *Schematic view of the Inner Detector with measured points originating from tracks produced in a Z^0 decay.*

MWPC layer for an isolated track varies from 0.5 to 1 mm depending on θ.

Since the start of the 1995 LEP running period, a longer ID has been operational, extending its acceptance in polar angle down to 15° and 165°. The MWPC layers have been replaced by 5 cylindrical layers of straw tube detectors measuring $R\phi$ and having the same functionality as the old MWPC trigger layers. The straw tubes provided no z measurement.

5.1.4 The Time Projection Chamber

The Time Projection Chamber (figure 5.5) is the main tracking device of DELPHI. It covers the active volume from $r = 35$ cm to 111 cm ($|z| \leq 134$ cm) and is filled by an argon-methane gas mixture (Ar/CH_4 : 80/20%). The TPC is divided into two hemispheres of six sectors in ϕ. It is read out at the end-caps by 16 concentric pad rows and 192 anode wires. In this way up to 16 three-dimensional space points can be measured per track. The dead space between the pads of adjacent end-plate sectors correspond to 4% of the $R\phi$ plane.

A track passing through the gas volume leaves a tube of ionisation along its way. Due to a homogeneous electric field along the z direction, the charges are drifted in the gas with a drift velocity of v_D=6.7 cm/μs to the end-caps. The $r\phi$ resolution is governed by the segmentation of the read-out structure, while the z coordinate is measured by the drift time. The single point resolution for tracks from multi-hadronic Z^0 decays is 250 μm in the $r\phi$ plane and 880 μm in the rz plane [DEL95]. The two-point separation is

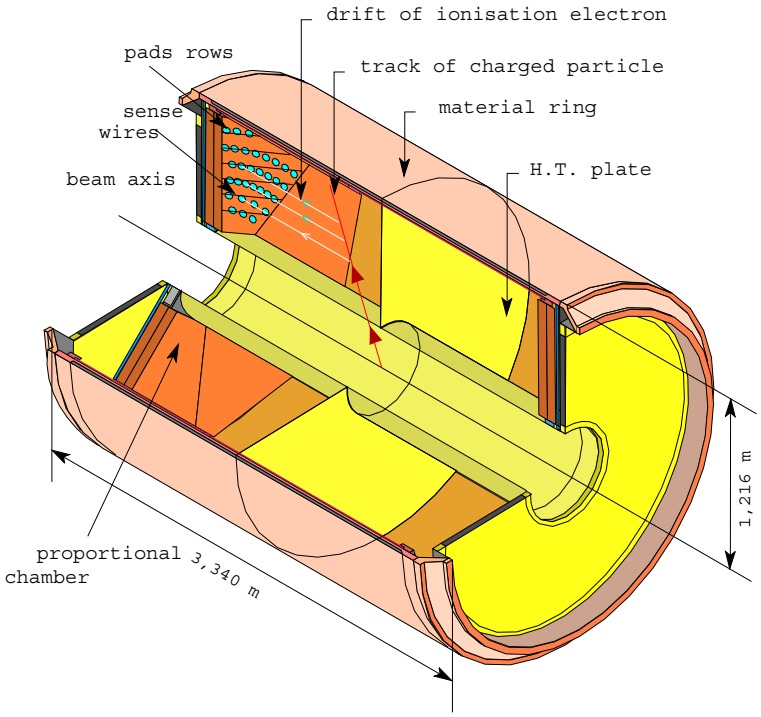

Figure 5.5: *Schematic overview of the TPC.*

of the order of 1 cm. Particle identification with the TPC is performed using the dE/dx measurement for charged tracks. This is described in detail in section 7.2.1.

5.1.5 The Outer Detector

The Outer Detector covers 4.7 m along the beam-axis at radii between 198 cm and 206 cm ($42° < \theta < 138°$). The OD consists of 24 modules, each of which has 5 layers of drift tubes operated in limited streamer mode. It is flushed with an admixture of argon (50%), isobuthane (48%) and isopropanol (2%). Each layer provides three-dimensional information. The resolution in $r\phi$ is 110 μm. The z information is obtained from the measurement of the arrival time of the signal at both ends, with a resolution of 3.5 cm.

The OD provides fast trigger information in both $r\phi$ and z direction. Since it is positioned $\approx$2 m away from the interaction point, the OD points increase the lever-arm for the track reconstruction and thus enable a precise determination of the momentum of a charged particle.

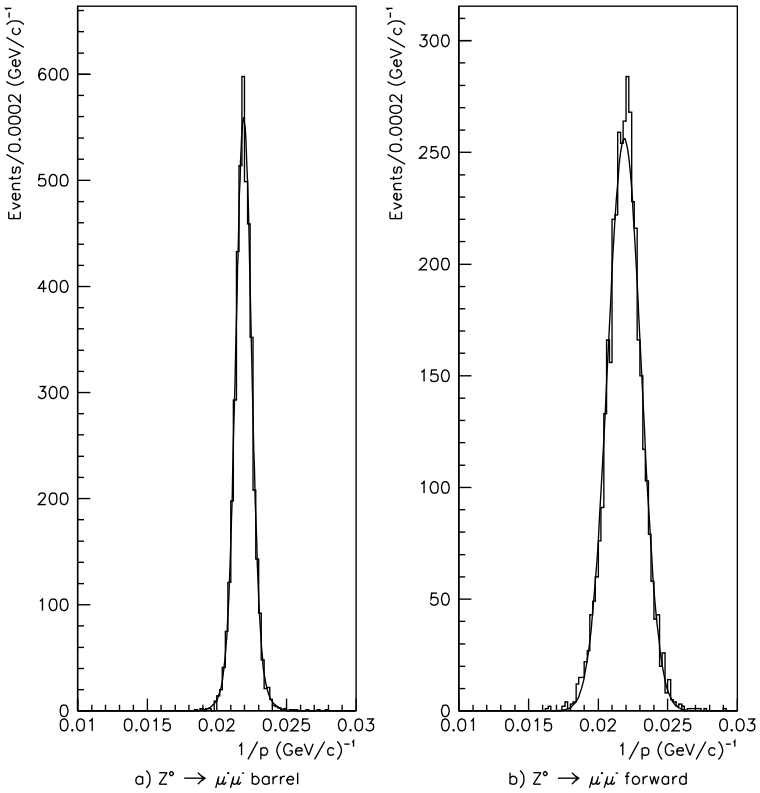

Figure 5.6: *Inverse momentum distributions for collinear muons from $Z^0 \to \mu^+\mu^-$ decays: (a) tracks containing hits from VD, ID, TPC and OD, (b) tracks containing hits from VD and FCB at least.*

5.1.6 Performance of the DELPHI Detector

It is important that all available tracking information is combined when reconstructing tracks in order to achieve the best possible momentum resolution. In particular, the performance of the RICH particle identification (Section 5.3) depends very much on the quality of the track reconstruction. DELPHI is designed such that, in the limited space available, one obtains both an optimal momentum resolution and an excellent particle identification.

In combining information from different detectors, it is of vital importance to have them globally well aligned. This is achieved by using 45.6 GeV/c muons from $Z^0 \to \mu^+\mu^-$. In the DELPHI barrel region a combination of track components obtained from the VD, ID, TPC and OD is used. Here, the total momentum resolution for muons from $Z^0 \to \mu^+\mu^-$ was measured to be (figure 5.6a):

$$\sigma(1/p) = 0.57 \times 10^{-3}\ (\mathrm{GeV/c})^{-1}. \tag{5.1}$$

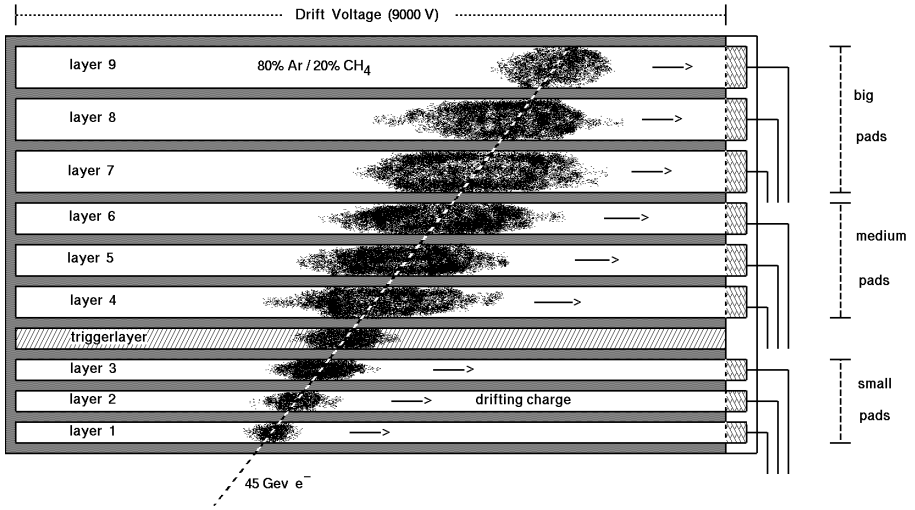

Figure 5.7: *Schematic view of the layer structure of a single HPC module. An entering electron initiates an electromagnetic shower in the converter. The produced charge is drifted to a planar MWPC. The full detector consists of 144 such modules.*

In the forward region for muons from $Z^0 \to \mu^+\mu^-$ with ($20° < \theta < 35°$) for which at least track information of the VD (inner layer) and the Forward Chamber B was available, the total resolution was measured to be (figure 5.6b):

$$\sigma(1/p) = 1.31 \times 10^{-3}\ (\mathrm{GeV/c})^{-1}. \tag{5.2}$$

5.2 Energy Measurement

5.2.1 The High Density Projection Chamber

The High density Projection Chamber (HPC) is the barrel electromagnetic calorimeter in DELPHI. It is the first large time-projection gas calorimeter, which provides a full three-dimensional reconstruction of an electromagnetic shower. It covers the angular region $42° < \Theta < 138°$. The HPC consists of 144 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 is a small TPC, with layers of a high density material in the gas volume (see fig. 5.7). These layers are made from lead wires which serve not only as converter material, but provide the drift field as well (E=106 V/cm). The total converter thickness is 18 $X_0/\sin\theta$. The lead gaps are filled with an argon-methane gas mixture (Ar/CH_4 : 80/20%).

The produced cloud of electric charge forms an electromagnetic shower. It is drifted with a velocity of v_D=5.5 cm/μs. The read-out of a single module is performed at the end of each module by a planar MWPC, which

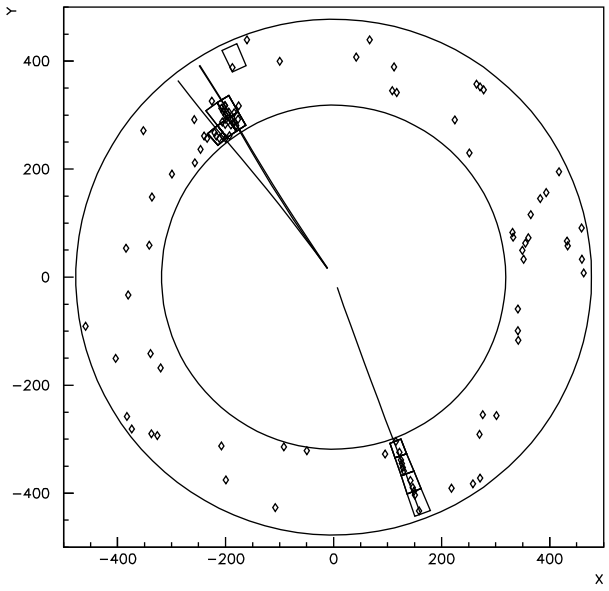

Figure 5.8: *A $\tau^+\tau^-$ event seen in the HCAL with cathode readout. One τ decays into three pions, the other one into a μ and two neutrinos (not seen).*

consists of 39 sense wires and is segmented in 128 pads. Each pad is read out in 256 time slots, providing very high granularity in z. This leads to a total number of $144 \times 128 \times 256 = 4.7 \cdot 10^6$ ADC signals which are available per event. The energy resolution of the HPC for photons has been found to be:

$$\frac{\sigma_E}{E} = \frac{32\%}{\sqrt{E[GeV]}} \oplus 4.3\% \ , \tag{5.3}$$

using neutral pions reconstructed with high precision from one photon converted before the TPC and one photon reconstructed in the HPC [DEL95]. The angular resolution for high energy photons is $\sigma_\phi = 1.7$ mrad in ϕ and in the polar angle $\sigma_\theta = 1.0$ mrad.

The θ resolution for 45 GeV electrons is constantly 0.6 mrad. The apparent ϕ resolution for electrons is 3.1 mrad, but much of this is due to the uncertainty in the electron track extrapolation from the TPC through the material of the RICH into the OD and finally to the HPC.

5.2.2 The Hadron Calorimeter

Both in the barrel and the forward regions, the return yoke of the superconducting magnet has been designed as an iron-gas Hadron Calorimeter (HCAL). The angular acceptance is between $43^\circ < \Theta < 137^\circ$ in the barrel region and between $11^\circ < \Theta < 50^\circ$ and $130^\circ < \Theta < 169^\circ$ in the forward

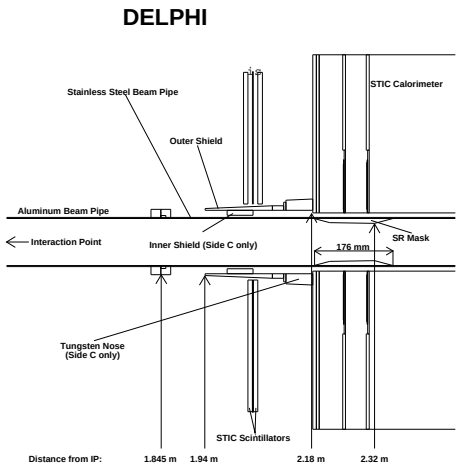

Figure 5.9: *Layout of the STIC region.*

regions.

More than 19,000 streamer tubes are installed in the 18 mm wide slots between the 50 mm thick iron plates in 20 layers in the barrel and in 19 in the forward regions. The chambers are filled with an admixture of CO_2 (59%), isobuthane (30%) and argon (11%).

For hadronic Z^0 events in the barrel region, the relative energy resolution in the HCAL is found to be:

$$\frac{\sigma_E}{E} = \frac{112\%}{\sqrt{E[GeV]}} \oplus 21\% \ , \tag{5.4}$$

where the fixed term is due to the material between the hadron calorimeter and the electromagnetic calorimeter. For calibration muons produced in $Z^0 \to \mu^+\mu^-$ decays are used. A typical event is presented in figure 5.8.

5.2.3 Luminosity Monitoring Detectors

At e^+e^- colliders luminosity is measured by counting the number of events of a process with a clear experimental signature, with high statistics and with a cross section which can be calculated theoretically with high precision. The process chosen is $e^+e^- \to e^+e^-$ Bhabha scattering at small angles, i.e. low Q^2, which proceeds almost entirely through the exchange of a photon in the t-channel.

In DELPHI, before 1994 the absolute luminosity was measured using the SAT detector and the VSAT detector was used to measure the relative luminosities at different energies.

The VSAT covers polar angles between $0.29^\circ < \Theta < 0.4^\circ$. It is a W-Si calorimeter of 24 radiation lengths giving a relative energy resolution of

$\approx 5\%$ for 45.6 GeV/c particles. In 1994 the SAT was replaced by a new calorimeter, the STIC (figure 5.9, built with the 'Schaschlik' technique. The STIC is a sampling lead-scintillator calorimeter, using wavelength shifting fibres and photoelectrode tubes for readout, placed on either side of the DELPHI interaction region at a distance of 2200 mm and covers an angular region between 29 and 185 mrad in θ.

Test beam measurements gave an energy resolution of $\sigma_E/E = 2.7\%$ for 45.6 GeV/c particles.

5.3 Particle Identification Devices

Powerful particle identification was made possible by the combination of dE/dx information from the TPC (and to a lesser extent from the VD) with information from the Ring Imaging CHerenkov counters (RICH) in both the forward and barrel regions. The RICH devices utilised both liquid and gas radiators in order to optimise coverage across a wide momentum range: the liquid radiator was used for the momentum range from 0.7 GeV/c to 8 GeV/c and the gas radiator for the range 2.5 GeV/c to 25 GeV/c.

This section emphasises the subdetectors used for kaon identification crucial for the spectroscopy of B_s^{**} mesons and to a lesser extent for $B_{u,d}^{**}$ mesons. The analysis of the detector measurements and their combination to provide an excellent identification tool is described in chapter 7.

5.3.1 Ring Imaging Cherenkov Detectors

The use of Cherenkov detectors made DELPHI unique among the four detectors installed at LEP. P.A.Cherenkov and S.I.Wawilow first showed that charged particles moving in a transparent medium may, under certain circumstances, produce electromagnetic radiation. This Cherenkov radiation was theoretically interpreted by I.E.Tamm and I.M.Frank.

Cherenkov Radiation

Electromagnetic radiation, of frequency ω, is produced in a transparent medium, when the velocity v of a charged particle exceeds the phase velocity of the emitted light in the medium concerned:

$$v > c/n(\omega),$$

where $n(\omega)$ is the refractive index and c is the velocity of light in vacuum. The radiation is emitted in a cone of opening angle $2\theta_c$ with

$$\cos\theta_c = c/nv = \beta/n = \frac{\sqrt{1 - m^2/E^2}}{n}. \tag{5.5}$$

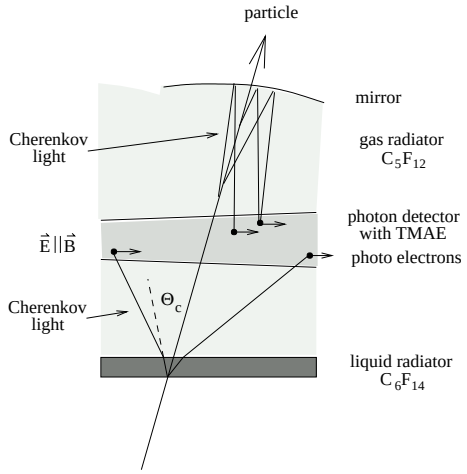

Figure 5.10: *The working principle of the Barrel RICH.*

The intensity of the radiation in a frequency interval $d\omega$ is proportional to $\sin^2\theta_c$. The number of photons in a given energy interval ΔE, for which the medium is transparent, is obtained by integrating over ω and over the time interval during which the particle traversed the medium of length L. For a particle of mass m and a given refractive index n there is a *threshold* velocity β_{th}, i.e. momentum $p_{th} = m\beta_{th}/sqrt(1-\beta_{th}^2)$, below which no radiation is emitted: $\beta_{th} = 1/n$. The emission angle is, according to 5.5, a function of the particle mass m and momentum p ($\approx E$ for $p \gg m$). For a given refractive index, the cone reaches a *saturated* angle, when the velocity of the particle is close to the speed of light in vacuum:

$$\cos\theta_c^{max} \approx 1/n \;\text{ for }\; \beta \simeq 1.$$

The Ring-Imaging Cherenkov Technique

The fact that the emission angle depends on the mass and momentum of the particle (5.5) is used to evaluate masses of charged particles. The main goal of the RICH detectors is 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 can also be used in the so-called *veto mode*.

The DELPHI RICH

In order to perform particle identification over a large momentum range of $0.7 < p < 45.6$ GeV/c, the DELPHI RICH contains two different radiator media. One is liquid perfluorhexane C_6F_{14} and the other gaseous perfluoropentane C_5F_{12} used in the barrel region. In the forward region C_4F_{10} is

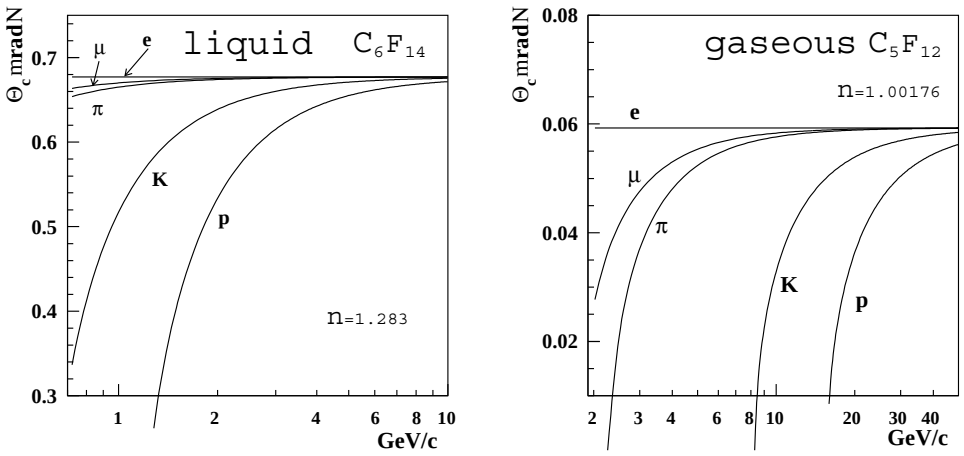

Figure 5.11: *The expected Cherenkov angle versus momentum for the liquid and gaseous radiators.*

used for the gas radiator. A schematic view of the Barrel RICH is given in figure 5.12.

The liquid radiator is used for particle identification in the momentum range from 0.7 to 4.0 GeV/c. The gas radiators are used for particles above 2.5 GeV/c. The expected Cherenkov angle versus momentum for the gas and liquid radiators is shown in figure 5.11. The Cherenkov light from the thin liquid radiator is projected onto a UV-sensitive photon detector, which converts UV-Photons into photoelectrons. Cherenkov light in the gas radiator is reflected to the same photon detector by focusing mirrors.

The UV-photon detector is a TPC-like device entirely made of quartz to allow UV-photon transmission from both sides. Inside the drift volume, the Cherenkov UV-photons will liberate electrons by ionisation of the drift gas, photosensitive TetrakisdiMethyl-Amino-Ethylene (TMAE), which is sensitive to photons of energies higher than $\sim$ 5.5 eV. A homogeneous electric field causes the electrons to drift towards the end of the photon detector, where they are detected by a Multi-Wire-Proportional-Chamber (MWPC).

The UV-photon detector provides three-dimensional information on the position of the UV-photon for each detected photoelectron. In the Barrel RICH, for instance, the MWPC gives $R\phi$ and R coordinates of the point where the Cherenkov photon liberated the photoelectron whereas the z coordinate of the point is obtained from the drift time measurement.

Particle Identification with the RICH

The potential of the RICH to identify particles is given by measuring the separation between the particle bands in figure 5.11 in units of the resolution

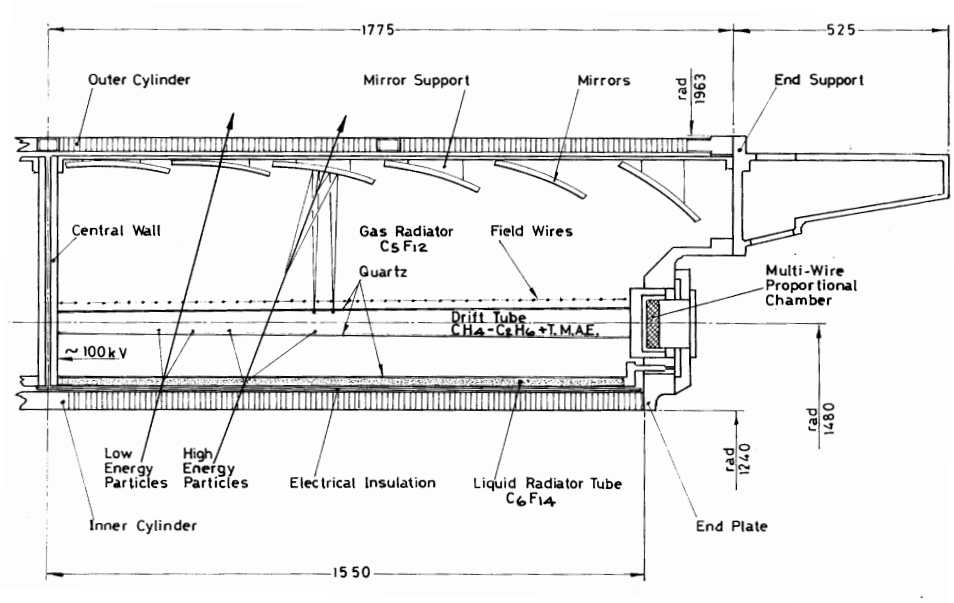

Figure 5.12: *Schematic view of the barrel RICH detector. The upper volume shows the gas radiator. The thin liquid radiator is indicated as well. The Cherenkov photons are converted in the TMAE gas. The converted electrons drift in both systems to the proportional chamber indicated on the right hand side.*

on the Cherenkov angle. The software performing the analysis of the data provided by the RICH used for particle identification is described in detail in section 7.2.1.

5.3.2 Specific Ionisation in the TPC

In addition to providing a three-dimensional track reconstruction, the DELPHI TPC is useful for charged particle identification by measuring the energy loss per unit length, dE/dx [DEL95]. The sense wires of its proportional chambers provide up to 192 ionisation measurements per track. The signals collected by the sense wires are associated to the tracks reconstructed by the TPC pads. This association is done by comparing the time of arrival of the pad and sense wire signals. Hits too close in time to be correctly separated are not used for the dE/dx calculation. This requirement corresponds, for tracks orthogonal to the drift direction z, to a separation of at least 2 cm.

On average of 5% of the signals collected by the sense wires are below the electronic threshold. The fraction of the Landau distribution lost due to this effect is a function of the drift length and gap size (i.e. wire spacing as seen from the track). The highest 20% of the signals measured are not used to reduce the influence of the Landau tail (due to abnormal energy loss or

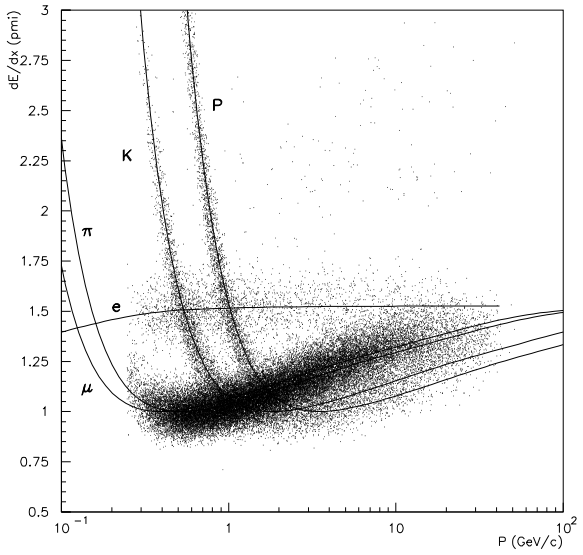

Figure 5.13: *Specific energy loss dE/dx in the TPC as a function of momentum. The full lines correspond to the expectations for e, μ, π, K and p tracks [DEL95].*

to δ rays).

To be used in the physics analysis, the truncated dE/dx is required to have at least 30 measurements. The efficiency obtained after these requirements is 61% for tracks in multihadronic events with momenta greater than 1 GeV/c and $|\cos\theta_{thrust}| < 0.7$. The dependence of dE/dx on the particle momentum p is measured from the data using various samples, and the final result can be seen in figure 5.13.

The resolution on dE/dx estimated from data is 6.7% for pions ($p >$ 2 GeV/c) coming from K^0_s decays in multihadronic events, and 5% for muons (p=45 GeV/c) from dimuon events ($Z^0 \to \mu^+\mu^-$). With the obtained resolution and dependence of dE/dx on p/m, the separation between electrons and pions is above 3 standard deviations for momenta below 4.5 GeV/c. A pion/kaon separation can be achieved at the 1σ level above 2 GeV/c.

Particle Identification with the TPC

The ability of the TPC to identify particles is dependent on the separation between the particle bands in figure 5.13 in units of the band width, in analogy to the particle identification with the RICH. The software performing the analysis of the data provided by the TPC and the RICH detectors used for particle identification is described in detail in section 7.2.1.

5.3.3 Specific Ionisation in the VD

The dE/dx measurement with the Vertex Detector is a part of the MAMMOTH [MF96] project, which was initiated to improve the track reconstruction of the DELPHI experiment.

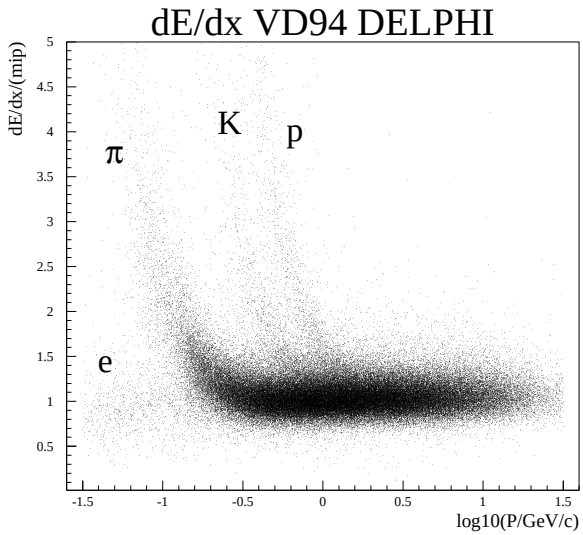

Figure 5.14: *The vertex detector dE/dx with three hits and more for the whole momentum range, measured with the 94 vertex detector. Clearly visible are the e, π,Kaon and proton bands.*

As in the TPC, a charged particle has a typical ionisation energy loss dE/dx in the vertex detector. Normally the VD is used for spatial measurements only, but the deposited signal in the silicon tracker gives also signal hight information. To calculate the dE/dx of a track in the vertex detector the associated hits are taken and the signal values are kept. After the path length correction and signal calibration, a maximum likelihood fit is applied and the Landau peak position calculated. For tracks with more than two hits the resulting dE/dx is shown in figure 5.14 for the whole possible momentum range of hadronic Z^0 decays. In this distribution several bands are visible, which allow a separation or particle identification. For the high momentum range the method is not competitive with the dE/dx measurement of the TPC. This comes from the worse resolution and the low relativistic rise of the dE/dx in silicon after the minimum. In the vertex detector only up to 6 points can be used while in the TPC a maximum of 192 wires give a signal. This explains the difference in the dE/dx resolution. For tracks with at least 3 VD hits the resolution is $\leq 15\%$.

Particle Identification with the VD

As for the TPC and the RICH detectors the separation between the particle bands in figure 5.14 in units of the band width provides the information for particle identification. The software combining the measurements from all three detectors, the VD, TPC and RICH, to build a powerful particle identification tool over the whole momentum range of tracks produced at LEP is described in detail in section 7.2.1.

5.3.4 Muon Chambers

The iron of the hadron calorimeter provides a filter which gives a first level of separation between muons and hadrons. Most hadrons are stopped by this material, whereas all muons of momenta above 2 GeV/c are expected to penetrate to the Muon Chambers.

The DELPHI muon chambers are drift chambers which are located behind the hadron calorimeter. They provide a muon identification with an efficiency of about 95%. The Barrel, Forward and Surround MUon (BMU, FMU and SMU) chambers cover the polar angular region between 15° and 165°. Resolution measurements on isolated tracks give $\sigma_{r\phi} = 4$ mm. The z coordinate is evaluated from delay time measurements (with a digitisation window of 2 ns) obtaining a resolution of $\sigma_z = 2.5$ cm [DEL95].

Charged particle tracks, reconstructed in the central detectors, are extrapolated through the solenoid. During this extrapolation the tracking errors are propagated and also the errors from multiple scattering are added. In both cases, correlations among the errors on the track parameters are taken into account. Thus, a precise measurement of the trajectories varies from 5-10 micrometers in the Vertex Detector, to a fraction of a millimetre in the Time Projection Chamber and to 1-3 mm in the Barrel Muon Chambers after traversing 5 m of the detector.

5.4 The DELPHI Analysis Chain

The recorded raw-data undergo a chain of treatments before they can be released for physics analyses. An overview of the chain is given in figure 5.15 in the form of a flowchart. It basically consists of three packages: an online software package to read out the raw data, a package for the simulation of events and of the detector response to enable the comparison of the real data with Monte Carlo data and an analysis package to prepare data for physics.

5.4.1 Trigger

The time window between two consecutive Beam Cross Overs (BCO) at LEP is 22 μs (11 μs in 8 bunch mode). The DELPHI trigger system is designed to handle high luminosities with a large background event rate and consists of four levels (T1, T2, T3 and T4).

The first two trigger levels are synchronous with respect to the BCO signal. The T1-decision is made after 3 μs, using ID and OD information and/or FCA and FCB information. In the barrel region the T1 is supported by the scintillating layer in the HPC. Within 39 μs after a positive T1 outcome (the T1-rate is of the order of 400 Hz), the T2 will have verified the information from the slow drift detectors (e.g. charged tracks in the TPC and electromagnetic showers in the HPC) to confirm the T1-decision.

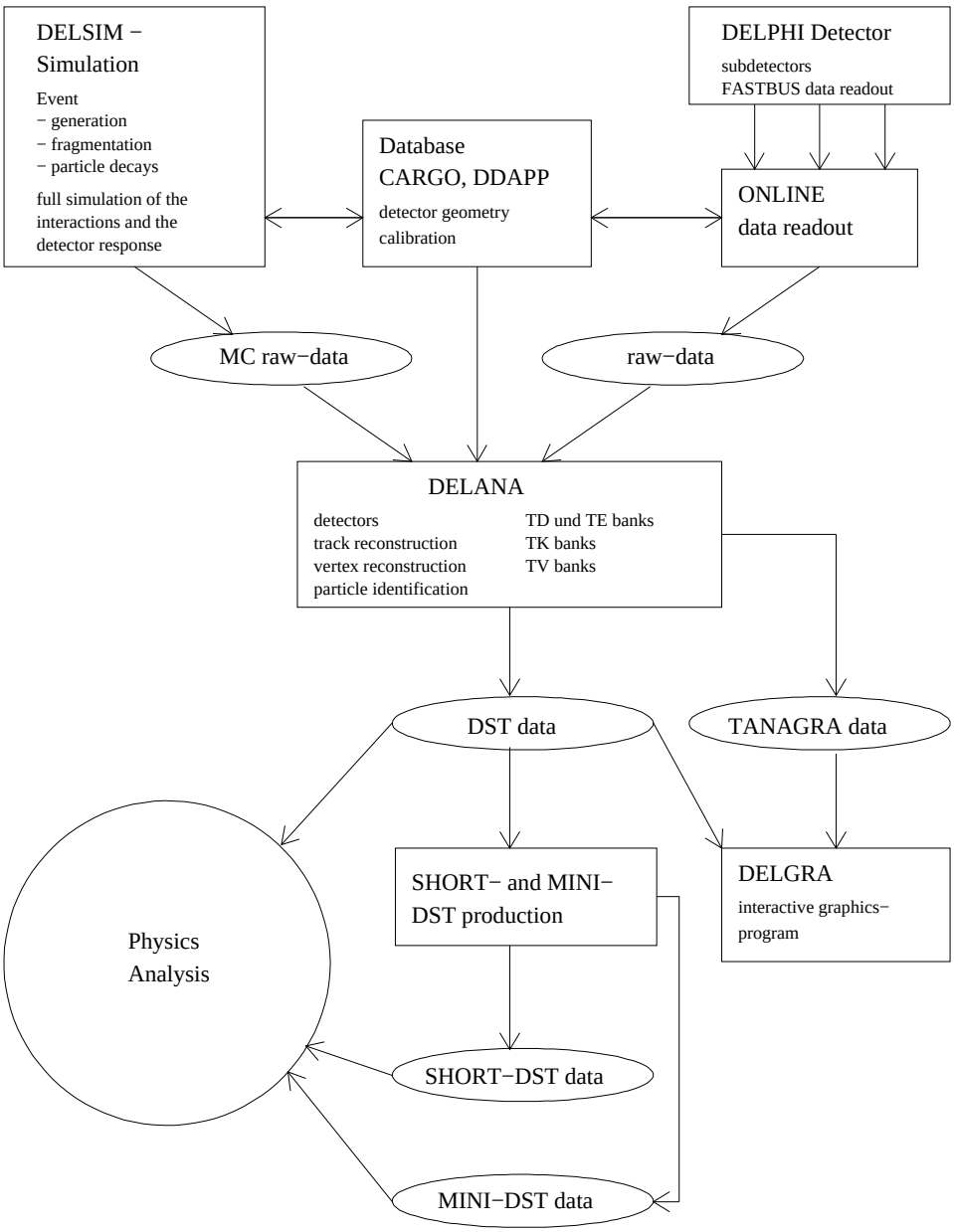

Figure 5.15: *The* DELPHI *analysis chain.*

The T2 rate is typically of the order of 4 Hz. The dead-time introduced is typically 3%, with 2% due to T1 and 1% to T2 for a typical readout time of 3 ms per event.

T3 and T4 are software filters performed asynchronously with respect to the BCO. The T2 rate is diminished to about 2 Hz by the T3, which uses the same logic but the full granularity and resolution to make its decision. The remaining background is further reduced by the T4 which flags events for physics analysis and simultaneously provides the information for the QC on the online event display. The trigger efficiency for multi-hadronic events is, due to their high final state multiplicity, almost 100% over nearly the full solid angle.

5.4.2 Data Acquisition System

The Data Acquisition System (DAS) stores all events triggered at the T2-level as *raw-data*. It uses the ZEBRA bank structure. ZEBRA is a memory management program for handling dynamic data structures. The size of a multi-hadronic event is of the order of 50 Kbyte. Each event is labelled with an individual number, the LEP fill number and the DELPHI run number. A run is defined as a time-period in which the detector was in stable conditions. Each run is recorded in a separate data file. Finally, the data-files are stored on IBM-3480 cassettes. At a later stage, discussed below, these tapes are read by the DELPHI offline system which is used for data-processing and physics analyses.

5.4.3 Slow Control System

The main task of the Slow Control System (SLOWC) system is to monitor and control the detector hardware and its electronics to ensure optimal detector stability for data-taking but also to guarantee maximum safety conditions.

The SLOWC system controls the switching of the low and high voltages of all detectors (*off*, *standby*, *ramping*, *ready* or *error*) as well as their gas systems, detector heating and cooling, pressure regulation systems and calibration systems. The hardware-connections are mostly provided by G64-micro-computers (90 in total) via Ethernet-links. The SLOWC basically comprises four inter-locked sub-systems: a display system for the status of the subdetectors, an error handling system, a gas control system and a separated safety system. Depending on the circumstances, both the safety and the error-handling system can take action without human intervention (the Big Brother System).

From the monitoring point of view, the SLOWC system surveys a large series of (sub)-detector parameters which may vary in time (e.g. temperatures, pressures, low- and high-voltages, drift velocities etc.) and stores the

information in the DELPHI-database. At a later stage, this information is folded with the recorded physics data.

5.4.4 Detector Simulation

The primary aim of a simulation program is to produce "data" for a particular reaction which are as close as possible to the real raw data from the detector. These "data" are then processed through the reconstruction program DELANA and the subsequent analysis programs in exactly the same way as for real data. This procedure models the detailed response of the complete detector to a particular physics subject. The DELPHI simulation (DELSIM) [Del89b] is based on three components which can be summarised as follows:

- A model for the generation of the primary physics process: In most cases, this is performed by external programs, like JETSET [Sjo94] and ARIADNE [Lon92] for quark final states. The DELPHI collaboration uses its own parameter tuning for the different generators determined from the latest available data i.e. lifetimes, braking ratios, fragmentation etc. This is a vital task for the quality of the simulated data, as it has to be compared with real data during an analysis.

- The general task of following particles through the DELPHI detector up to the point where they hit an active detector component: This is done by stepping through the magnetic field including the possibility that these particles give rise to secondary interactions. In order to reach the required accuracy for the results, a very detailed description of the material inside the detector is necessary.

- Following the particles inside the active detector components and the simulation of the detector response as recorded in reality: This part is specific to every detector component and the modularity of the code is such that each of them corresponds to an independent software module.

The particle tracking through the DELPHI detector includes energy loss, multiple scattering and the following secondary processes:

- the photoelectric effect
- the emission of δ rays
- bremsstrahlung
- annihilation of positrons
- pair production

- Compton scattering
- weak decays
- nuclear interactions

The material parameters which determine the rates of the above processes are extracted from the CARGO data base. When a particle enters an active detector component the control of the track-following is given to the corresponding software module. Most modules follow the particles by using tools provided by the general routines outlined above. When a particle crosses the sensitive volume of a detector the relevant information is stored to compute the detector response in the form of electronic signals, as for real data.

In addition to this detailed simulation a simplified version, called FASTSIM [CMR+89], has been implemented to obtain a much faster simulation of the DELPHI events. This version uses a simplified description of the material in the detector and calls detector modules which simulate directly the results of the pattern recognition, thus avoiding the time consuming step of producing the raw data.

The simulation of events for physics analyses and detector studies is produced in several production centres in the DELPHI member states. The achieved statistics is comparable with or larger then the statistics for the actual LEP data.

5.4.5 Reconstruction Program

The main reconstruction program is the DELphi ANAlysis program DELANA [Del89a]. It contains one module for each subdetector which performs a necessary alignment and calibration of the raw data. The data format inside DELANA is based on ZEBRA. The backbone of DELANA is the Track ANAlysis and GRAphics package TANAGRA which provides a well defined data structure for storing track and vertex information in a format independent of the various program modules.

The event reconstruction proceeds in the following steps:

- A local pattern recognition is performed independently and separately for each subdetector resulting in so-called Track-Elements (TE) e.g. space points and directions, energy depositions and so on. This step is known as *first stage pattern recognition*.
- The track elements from several detectors are grouped into track candidates and a first track fit is performed. The track fit accounts for multiple scattering and energy loss in the material between the measurements.
- After resolving ambiguities, the tracks are extrapolated to obtain precise estimates for their passage through the subdetectors.

- A *second stage pattern recognition* is performed, where the local pattern recognition is redone using the appropriate extrapolated information from the other subdetectors.
- Energy depositions in the calorimeters are either successfully linked to charged tracks or are marked as neutral energy.
- Primary event vertices are fitted from the reconstructed tracks.

The output of this procedure is the so-called Data Summary Tape (DST) [Sac94] format containing a tree structure of banks based on the reconstructed vertices. All information related to physics analysis (e.g. four-vectors of particles, calorimeter information, matches to the simulation etc.) is stored here. The average size of a multi-hadronic Z^0 DST event is roughly 60 Kbyte.

5.4.6 SDST production

In order to improve data quality and to provide reasonable particle identification for the collaboration, an additional processing is performed. It reduces the amount of data by a factor of three and is called Short DST (SDST). It contains the following:

- The tagging of $Z^0 \to b\bar{b}$ events.
- e, γ and π^0 identification.
- $K^{\pm}$ and p identification.
- Muon identification.
- Reconstruction of V^0 and Λ^0 decays.
- Reconstruction of VD only tracks.
- Reconstruction of interactions with the detector material.

Chapter 6

Artificial Neural Networks

Artificial Neural Networks (NN) have seen an explosion of interest over the last few years. NN are being successfully applied across an extraordinary range of problem domains, in areas as diverse as finance, medicine, engineering, geology and physics. Their success can be attributed among other things to the modelling of non-linear complex functions, a domain where the otherwise well understood linear models are not valid.

A description of neural networks can be found in various textbooks ([Bis95], [Car92], [Fau94], [Hay94], [Pat96], [Rip96]) and a very good summary at the URL [Sta]. The discussed features are, among many others, realized in the program package NeuroBayes of the company PhI-T [PT], a physics spin-off with contributions from this and many similar analyses of the Karlsruhe DELPHI group.

6.1 Introduction

To copy the ability of the brain to learn patterns, i.e. to produce an appropriate reaction to a visual or acoustical input, is the aim of transferring the basic working principle of biological neural systems to artificial neural networks.

The wanted feature of a NN is to produce an output y when presented with an input $\mathbf{x}$. The input vector $\mathbf{x}$ contains, for the case of particle identification, variables describing measurements of the specific ionization energy loss in the TPC and the vertex detector and the information gathered in the RICH detectors (see chapter 5). The desired output of the network t is a decision if the particle is a kaon or not. The output $t = +1$ standing for a kaon and the output $t = -1$ for a non-kaon.

A NN consists of layers of inter-connected nodes. The topology of nodes used for particle identification is arranged in three layers. The first layer of nodes is directly connected to the input data, the second layer called the hidden layer is connected to the first layer with weights controlling the

strength of the coupling between nodes. The input of a node is the linear combination $y = \sum w_i x_i$ of all input layer nodes. The output of a hidden node is a function $g(y)$ of its input, where g is called the *activation function*. The third layer consists of just one node, the output node. It is connected to all hidden nodes. Its output $g(y')$ is a function of y', which is the linear combination of outputs of the hidden layer nodes.

The actual output of the network y is thus dependent on the input data and the adaptive weights connecting the nodes, the artificial neurons. For each input pattern there will be a difference between the desired output t and the actual output y. The sum of the residuals $E_n = y_n - t_n$ is the error made by the network using the given weights. The learning procedure of a neural network consists in adjusting these weights such as to minimize the error. Since the error is dependent on the weights, the error surface is a multi-dimensioned mountain. The aim of network learning is to find the deepest valley in the weight space. For a given set of weights the direction which reduces the error the most is given by the negative gradient, just like finding the valley between two mountains can be most easily achieved by walking downhill. The method used for network training is thus called gradient descent.

The performance of the network prediction of the particle type is dependent on its generalization ability. The quality of particle identification is not given by the size of the error at the end of network training but by the ability of assigning a previously unseen particle to the correct class. The number of weights and nodes building up the network has to be as large as necessary to be able to adjust the underlying features of particles but also small enough in order not to over fit the data.

In addition to class assignment NN can also be used for predicting complete probability density functions (PDF). This feature is used for example for the measurement of the B_s^{**} mass. The output is not just a continuous variable but a complete function giving the size of belief of the true position of the mass value. The PDF can be used to evaluate the most probable value or the very important information about the error of the measurement by considering the width of the distribution.

6.2 The Biological Inspiration

Research in the area of Artificial Intelligence has led to the perception that in order to reproduce intelligence, it would be necessary to mimic the underlying structure of the brain.

The brain is composed of approximately 10^{10} neurons, massively interconnected. Each neuron is a specialized cell which can propagate an electrochemical signal. A neuron has a branching input structure, a cell body and a branching output structure. The output of one cell connects to the

input of another via a gap, called the synapse.The addressed neuron fires an output signal only if the total signal received via the inputs exceeds a certain level, the firing threshold. Learning consists of altering the strength of synaptic connections.

A crude model of the brain and its working principle has been successfully implemented in artificial neural networks to achieve some remarkable results. Nevertheless, the complex functionality of the brain is not achieved by any artificial neural network and a comparison is not possible beyond the very low-level structure.

6.3 The Basic Artificial Model

This section describes the basic idea, how artificial algorithms can incorporate functions of biological systems. Signal processing units introduced here can be integrated to complex neural networks capable of predicting the parameters of non-linear problems.

The basic constituents of artificial neural networks are nodes, corresponding to biological neurons, and weighted connections, corresponding to the synapses, controlling the strength of node coupling. Each input x_i to a node has a strength or weight, w_i. In addition, a node also has a single threshold value w_0, called the bias[1]. The role of the bias will get clear at the end of this section. The weighted sum of the inputs is formed and the threshold added to compose the test statistic of the node

$$a(\mathbf{x}) = \sum_i w_i x_i + w_0. \tag{6.1}$$

The output y of the neuron is given by the *activation function* $y(x) = g(a(x))$. It is generally chosen to be monotonic. The activation function can be the step function giving the output 0 if $a \leq 0$ and the output 1 if $a > 0$. A more general choice of activation function is the *sigmoid* function

$$g = \frac{1}{1 + e^{-a(\mathbf{x})}} \tag{6.2}$$

shown in figure 6.1. The motivation for this form of discriminant follows from Bayes' theorem, where the posterior probability of membership of class $\mathcal{C}_1$ for a two-class problem is given by

$$\begin{aligned} P(\mathcal{C}_1|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1) + p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2)} && (6.3)\\ &= \frac{1}{1 + e^{-a(\mathbf{x})}} && (6.4)\\ &= g(a(\mathbf{x})) && (6.5) \end{aligned}$$

[1]The bias introduced here must not be mistaken with the bias defined in statistics describing the residual between the true and the estimated value of a parameter.

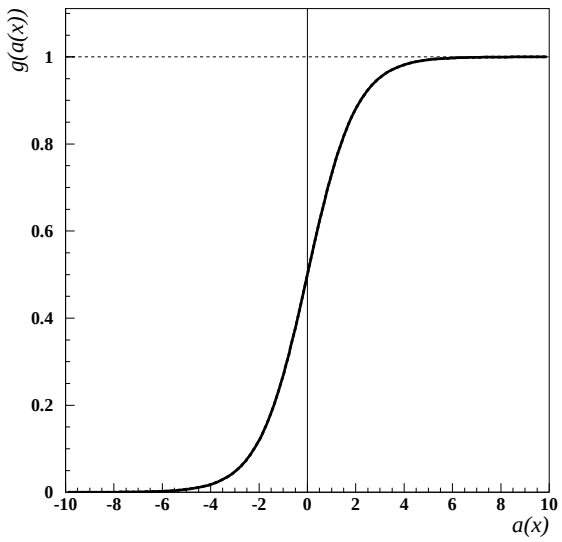

Figure 6.1: *The sigmoid activation function.*

where

$$a(\mathbf{x}) = \ln \frac{p(\mathbf{x}|\mathcal{C}_1)P(\mathcal{C}_1)}{p(\mathbf{x}|\mathcal{C}_2)P(\mathcal{C}_2)}. \tag{6.6}$$

If the class-conditional probabilities $p(\mathbf{x}|\mathcal{C}_i)$, i.e. the likelihood functions are given by Gaussian distributions with equal covariance matrices, $a(\mathbf{x})$ takes the form of equation 6.1. More precisely, if the likelihood functions are normally distributed *and* the sigmoid activation function is used, the output of the network can be interpreted as a posterior probability.

The term sigmoid means 'S-shaped', and the logistic form of the sigmoid maps the interval $(-\infty, \infty)$ onto $(0, 1)$. Due to the non-linear mapping performed by the sigmoid function, the output neuron is only sensitive to the small range around zero. Values lying outside of this region are 'squashed' by the transformation, i.e. there is a saturation effect. The effect of the bias is to shift the mean of the input data distribution to the approximately linear area of the sigmoid. In other words, the bias is compensating for the difference between the mean value of the target distribution and that of the input distribution.

If $|a|$ is small, the logistic sigmoid function $g(a)$ can be approximated by a linear function, and so in this sense a network with sigmoidal activation functions contains a linear network as a special case. For the extreme case that all weights are set to zero the network prediction is equal to the á priori knowledge. For example, for a two class problem with equal class sizes $P(\mathcal{C}_1) = P(\mathcal{C}_2)$ the constant network prediction, for vanishing weights,

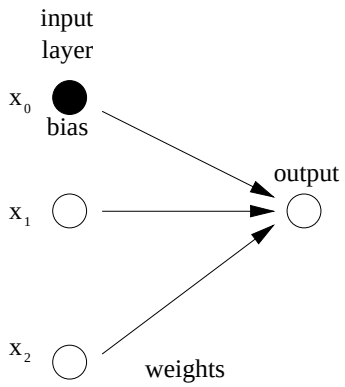

Figure 6.2: *Representation of a linear discriminant function as a neural network.*

is $g(0) = 0.5$, which reflects the probability to extract a member of a given class from the complete data set.

For the application of neural networks in this analysis, it is more convenient to use a transformed version of the sigmoid. The used function is obtained by $g'(a) = 2g(a) - 1$. This transformation results in the output -1 for background and $+1$ for signal, but more importantly, it results in the output 0, if there is no information provided to the network by the input data. This is a more natural result than 0.5. Connections to consecutive layers of nodes are so automatically set to zero.

6.4 The Perceptron

With the basic concept of data processing on the level of one node introduced above the structure of neural networks can be built up. NN with their adjustable weights represent *perceptrons* capable of learning the features of the input data.

The realization of the inputs connected to the output via adjustable strengths can be regarded as a form of neural network with a single layer of adaptive weights between the inputs and the outputs, and is called a *single-layer perceptron*. The threshold w_0 introduced in (6.1) can be incorporated into the network. By defining $\tilde{\mathbf{x}} = (1, \mathbf{x})$ and $\tilde{\mathbf{w}} = (w_0, \mathbf{w})$, (6.1) can be written in the form

$$a(\mathbf{x}) = \tilde{\mathbf{w}}^{\mathrm{T}} \tilde{\mathbf{x}}. \tag{6.7}$$

The bias w_0 is represented as a weight from an extra input x_0 which is permanently set to unity.

Single layer perceptrons are capable of separating problems by drawing a decision boundary between members of different classes. In a problem with two input variables the decision boundary is a straight line as can be

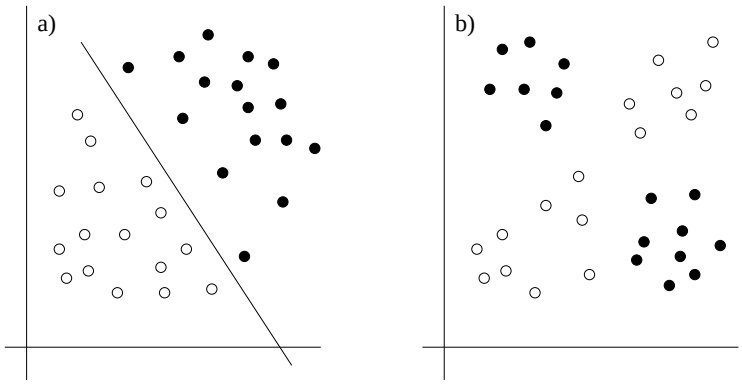

Figure 6.3: *Demonstration of linear separability and non-separability. In a) a linear decision boundary can perfectly separate the members of the two classes. In b) no such boundary can be drawn without including members of the other class. This problem can only be solved with neural networks containing at least two layers of weights.*

seen in figure 6.3. If the problem is linearly separable the classification of all cases can be done correctly. In a linearly non-separable problem no linear decision boundary can be drawn, which could perfectly assign all patterns to their corresponding classes.

Typical problems in high energy physics are not linearly separable since distributions belonging to various classes overlap. Particle identification, B_s^{**} mass reconstruction and the correction on the weak decaying B direction are based on neural networks capable of separating non-linear problems.

To allow for non-linear mappings, networks with several layers of adaptive weights can be introduced. In fact, neural networks with just two layers of weights are capable of approximating *any* continuous functional mapping[2]. Nodes of networks used in this analysis rely only on nodes of the previous layer and not on the output of later layers. The data is 'fed' in at the input layer, multiplied by the weights connecting to the first hidden layer to give their input. With the inputs given, the activation function is applied for each hidden node. Their outputs are transfered via the second layer of weights to give the input to the second hidden layer of nodes. This is continued until the output layer of node(s) is reached. This process corresponds to a forward flow of information. This network type is thus called *feed-forward*.

This concept ensures that the network outputs can be calculated as explicit functions of the inputs and the weights. Such networks are generally

[2]This has been formally proven by Kolgomorov [Kol57], who has shown that every continuous function of several variables can be represented as the superposition of a small number of functions of one variable.

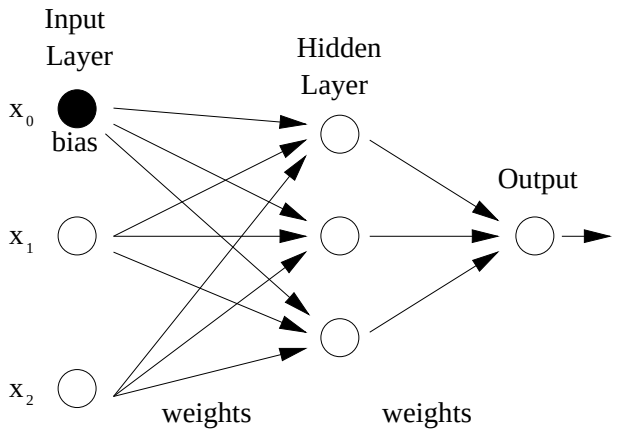

Figure 6.4: *Example for a multilayer perceptron.*

called *multi-layer perceptrons*. The complete function represented by a 2-layer network (see figure 6.4) is given by the explicit expression

$$y = \tilde{g}\left(\sum_{j=0}^{M} w_j^{(2)} g\left(\sum_{i=0}^{d} w_{ji}^{(1)} x_i\right)\right) \tag{6.8}$$

where d is the number of input units and M is the number of hidden units.

In practice, the number of input and output units is defined by the problem. The number of hidden layers and units to use is far from clear. It depends on the correlation between the input variables and must be adjusted to the special case considered.

6.5 Training a Neural Network

So far, the general structure of neural networks has been discussed. The remainder of this chapter will deal with techniques for the training of such networks. Training a network is the process of changing the weights so that the residual between the output and the target is minimized.

The output of the network for a given input pattern x_n is $y_n = y(x_n, \mathbf{w})$, which is evaluated by the feed-forward technique described in the last section. The target t_n corresponding to x_n is the desired output of the network. The difference between the actual and the desired network output is the error made by using the actual set of weights.

The first important point is the choice of error function. From the many existing possibilities for the present analysis only the two following are relevant. The sum-of-squares error and the entropy error function. The first one is mainly used for its simplicity and the latter for its advantageous characteristic when trained with an unbalanced amount of data classes. Both

error functions allow the output of the network to be interpreted as a posterior probability, which is of great importance for the case of reconstructing probability density functions.

The second important step of network training is the error minimization process itself. In the parameter space of the weight, the error function represents a multi-dimensional 'mountain' with many peeks and valleys. The task of network training is to adjust the weights in order to move step-by-step towards the global minimum, i.e. the deepest valley of the weight space. The process of error minimization is performed by the combined method of gradient descent and back-propagation. Since the weight-space is extremely ragged, the training process can be extremely difficult, time-consuming or even impossible. Simple but helpful methods can be applied to ensure stable network training and to allow for finding a better error minimum.

By applying these methods, which will be described in detail in the following, network training has become considerably easier. The training time has been reduced by at least of an order of magnitude by improving at the same time the network prediction quality. The reconstruction of the B_s^{**} mass and of the B direction have considerably improved the Q-value resolution, which is essential to the analysis.

6.5.1 Error Functions

Standard training of neural networks minimizes the square of the error function

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(\mathbf{x}_n, \mathbf{w}) - t_n\}^2 . \tag{6.9}$$

called the sum-of-squares error function, where N represents the number of training patterns. Since $y^n = y(\mathbf{x}^n, \mathbf{w})$ is a linear function of the weights, the error function $E(\mathbf{w})$ is a quadratic function of the weights, and hence its derivatives with respect to the weights are linear functions of the weights. This linearity of the derivatives is the most important feature of the sum-of-squares error function, besides for its analytical simplicity. The system of equations given by the derivatives represents a set of coupled equations known as the normal equations of the least-square problem, for which an analytic solution can be given. This solution results in one step in the estimated values for the network weights.

The sum-of-squares error function has been used for the training of the kaon and proton identification networks, described in the next chapter. There are also other choices of error functions, but if the output of the neural network is to be interpreted as a probability, the error function has to satisfy the condition

$$\frac{E'(1-y)}{E'(y)} = \frac{1-y}{y}. \tag{6.10}$$

A class of functions E, which satisfies this condition, is given by

$$E(y) = \int y^r (1-y)^{r-1}\, dy. \tag{6.11}$$

This includes two important error functions. For $r = 1$

$$E(y) = \frac{y^2}{2} \tag{6.12}$$

which gives the already mentioned sum-of-squares error function. For $r = 0$

$$E(y) = -\ln(1-y) = -\ln(1-|y|) \tag{6.13}$$

which gives rise to the entropy error function. To see this, consider that $E(y,t) = -\ln(1-|y-t|) = -\ln(y)$ if the target $t = 1$ and $E(y,t) = -\ln(1-|y-t|) = -\ln(1-y)$ if the target $t = 0$. Summing over all training patterns, this can be combined into an expression of the form

$$E(\mathbf{w}) = -\sum_{n=1}^{N} \{t_n \ln y(\mathbf{x}_n) + (1-t_n)\ln(1-y(\mathbf{x}_n))\}\,. \tag{6.14}$$

The entropy error function is preferable if there is an unbalanced amount of members in different classes. This is the case for almost all considered physics problems, since neural networks are only applied for cases with a small signal rate. The $B^{**}_{u,d,s}$ networks described in the next chapter are trained using the entropy error function. The signal to background ratio for the case of the B^{**}_s network is of the order of 10^{-3} at the track level. The á priori network output would be 0, zero standing for background, since only in the very seldom case of $1:10^3$ a prediction error would occur. If the sum-of-squares error function is used the error contribution of the single signal event in a thousand background events would not be significant enough to force the weights of the network to move away from this starting point. On the other hand, if the entropy error function is used, for a completely wrong classification, the error function gives an infinitely large contribution. For a true B^{**}_s event the target value is $t = 1$. In the above equation the second term vanishes, but the first term gives, with the wrong prediction $y = 0$, $E_n = \infty$. In this case the network would be forced to adapt to the seldom, but existing, signal events in the data in order to minimize the prediction error.

6.5.2 Gradient Descent

For a linear network, the weight values minimizing the sum-of-squares error function can be explicitly found in terms of a matrix equation. If a non-linear activation function, such as a sigmoid, is used or if a different error function

is considered, a closed form solution is no longer possible. This mainly means that the solution of the normal equations introduced in the last section does not directly, i.e. in one Newton step, lead to the global minimum of the error minimization problem. It rather provides a method for slowly reducing the error step-by-step. This procedure can be geometrically interpreted as a descent along the steepest direction. This direction is given by the negative of the gradient at a given point. The method is thus called *gradient descent*.

The only condition for the usage of gradient descent is that the activation function has to be differentiable, as it is for the case of the sigmoid. With an initial guess for $\mathbf{w}$, the update of the weight vector is given by moving a small distance in $\mathbf{w}$-space in the direction in which E decreases most rapidly, i.e. in the direction of $-\nabla_{\mathbf{w}} E$. By iterating this, process a sequence of weight vectors $\mathbf{w}^{(\tau)}$ is generated whose new components $w_j^{(\tau+1)}$ are calculated by updating the old values $w_j^{(\tau)}$

$$w_j^{(\tau+1)} = w_j^{(\tau)} - \eta \left. \frac{\partial E^n}{\partial w_j} \right|_{\mathbf{w}^{(\tau)}} \tag{6.15}$$

where η is a small positive number called the *learning rate* parameter. Under suitable conditions the sequence of weight vectors will converge to a point at which E is minimized. The choice of the value for η can be fairly critical, since if it is too small the reduction in error will be very slow, whereas, if it is too large, divergent oscillations can result. For a network with a differentiable activation function, the derivatives are given by

$$\frac{\partial E^n}{\partial w_j} = g'(a)\{y(\mathbf{x}^n, \mathbf{w}) - t^n\}x^n = g'(a)\delta^n x^n \tag{6.16}$$

with $\delta^n \equiv y(\mathbf{x}^n, \mathbf{w}) - t^n$. Combining (6.16) and (6.15), the change in the weights due to presentation of a particular pattern is given by the *delta rule*

$$\Delta w_j = w_j^{(\tau+1)} - w_j^{(\tau)} = -\eta\, g(a)(1 - g(a))\, \delta^n x^n. \tag{6.17}$$

since for the sigmoid, the derivative of the activation function can be expressed in the simple form $g'(a) = g(a)(1 - g(a))$.

6.5.3 Error Back-Propagation

The method of minimizing the error function by gradient descent, can be, in the form presented above, only applied to single-layer networks, since it is dependent on the residual $\delta = y - t$. The reason is that for hidden layers the target output of the nodes is not defined. Back-propagation describes the method of tracing back the source of the error made by the node in the output layer by propagating the error backwards through the network, by applying the chain rule for partial derivatives. This way the derivatives of

the error function can be evaluated with respect to any weight parameter regardless of its position within the network. With the derivatives given, the usual gradient descent method can be applied to compute the adjustments to be made to the weights.

In a general feed-forward network, each unit computes a weighted sum of its inputs to the form

$$a_j = \sum_i w_{ji} x_i \tag{6.18}$$

where x_i can be the activation of a previous unit, or an input and w_{ji} is the weight associated with that connection. The sum is transformed by a non-linear activation function to give the activation x_j of unit j in the form

$$y_j = g(a_j). \tag{6.19}$$

The network error E_n can be expressed as $E_n = E_n(y)$. For each pattern, the input vector is supplied to the network and the activations of all units are calculated layer-by-layer. This *forward propagation* can be regarded as a forward flow of information through the network.

The derivative of E_n with respect to some weight w_{ji} is given by

$$\frac{\partial E^n}{\partial w_{ji}} = \frac{\partial E^n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}} \tag{6.20}$$

by applying the chain rule. The second derivative on the right-hand side is given by

$$\frac{\partial a_j}{\partial w_{ji}} = x_i. \tag{6.21}$$

With this result, the derivative of E_n becomes

$$\frac{\partial E_n}{\partial w_{ji}} = \delta_j z_i. \tag{6.22}$$

where δ_j is the *error* defined by $\delta_j = \partial E^n / \partial a_j$. This formula has the same general form as the derivative for single layer networks. To obtain the required derivative, the value of δ_j for the unit at the output end of the weight has to be multiplied by the value of x_i for the unit at the input end of the weight.

The evaluation of δ for the output unit is straightforward

$$\delta_k \equiv \frac{\partial E^n}{\partial a_k} = \frac{\partial E^n}{\partial y_k} g'(a_k) \tag{6.23}$$

with $y_k = g(a_k)$. The δ's for hidden units are given by

$$\delta_j \equiv \frac{\partial E^n}{\partial a_j} = \sum_k \frac{\partial E^n}{\partial a_k} \frac{\partial a_k}{\partial a_j} \tag{6.24}$$

where the sum runs over all units k to which unit j sends connections. This formula makes use of the fact that variations in a_j give rise to variations in the error function only through variations in the variable a_k. Substituting the results into the definition of δ gives the *back-propagation* formula

$$\delta_j = g(a_j)(1 - g(a_j) \sum_k w_{kj}\delta_k \tag{6.25}$$

which says that the value of δ for a given unit can be obtained by propagating the δ's backwards from units higher up in the network. Since the value of δ for the unit is known, it follows that by recursively applying the back-propagation formula, the δ's can be evaluated for all of the hidden units in a feed-forward network.

In order to turn this into a learning algorithm, the weight updates have to be based on these derivatives. Considering the fixed-step gradient descent technique introduced above for the single-layer network, the weight updates for the first layer become for *batch learning*, i.e. a weight update after N patterns,

$$\Delta w_{ji} = -\eta \sum_{n=1}^{N} \delta_j^n x_i^n \tag{6.26}$$

with analogous expressions for the weights of later layers.

6.5.4 The Hessian Matrix

Back-propagation can also be used to evaluate the second derivatives of the error, given by

$$H_{jk} = \frac{\partial^2 E}{\partial w_{ji} w_{lk}}. \tag{6.27}$$

These derivatives form the elements of the *Hessian matrix* and are used to estimate the optimal learn rate in the multi-dimensional variable space. which plays an important role in the training of the neural networks applied in this analysis. The significance of input variables with respect to the network output can be evaluated by the inverse of the Hessian Matrix. The eigenvalues of this matrix thus represent the inverse of the curvature at a given point. This is in analogy to the variance of maximum likelihood estimators, which can be estimated by the inverse of the second derivative of the likelihood function, i.e. the inverse of the curvature. A well measured parameter has a 'sharp' likelihood function and a poor measurement gives a shallow likelihood function. Since different components of the input vector have different significance with respect to the network output, the parameter valley of the weight space can be shaped canyon like. To one direction there are almost vertical walls, to an other it is completely flat. A common learning parameter η from equation (6.26) can make network learning very unstable or very slow. The Hessian Matrix can not only be used for determining the

shape of the error valley but also to modify the step direction given by the gradient descent method to optimize network training. How this has been achieved for the current analysis is described in section 6.8.

The exact calculation of the Hessian Matrix has proven to be impractical for the applied networks, since the number of function evaluations scales with N_w^2, with N_w being the number of weights in a network. A typical value for N_w is 1000. The Hessian is approximated by its diagonal since its inverse is trivial to evaluate. Doing so, the correlations between input variables are neglected.

Even for uncorrelated input variables, in which case the Hessian is by definition diagonal, the elements of the matrix are not explicitely calculated but are approximated. There is a way of calculating the product of the Hessian with a vector without having to compute the Hessian itself. The finite difference method can fulfil this task for an arbitrary vector $\mathbf{v}$

$$H\mathbf{v} \approx \frac{1}{\alpha}\left(\frac{\partial E}{\partial w}(w+\alpha\mathbf{v}) - \frac{\partial E}{\partial w}(w)\right) \tag{6.28}$$

using the gradient at the points w and $w+\alpha\mathbf{v}$ respectively. This method can be applied to compute the eigenvectors and eigenvalues of H by the *power method*. By iterating and setting

$$\mathbf{v}(t+1) = \frac{H\mathbf{v}(t)}{||\mathbf{v}(t)||} \tag{6.29}$$

the vector $\mathbf{v}(t)$ will converge to the largest eigenvector of H and $||\mathbf{v}(t)||$ to the corresponding eigenvalue. Eigenvectors corresponding to smaller eigenvalues can be then calculated by setting the elements of $\mathbf{v}$ corresponding to larger eigenvalues to zero after each iteration. This way, every eigenvector and eigenvalue of the Hessian Matrix can be approximately calculated.

6.6 Prediction Problems

6.6.1 Classification

In classification, the general objective is to determine if an input pattern belongs to class a or b, when the collection of all patterns belongs to the class $a+b$. The network output is, for such a two-class problem, a value between -1 and $+1$. The closer the network output is to $+1$, the probability of the pattern to belong to class a increases, whereas if the output gets close to -1 the probability of the pattern to belong to class b increases.

The MACRIB package, extensively used for kaon identification throughout this analysis, provides such a classification decision. The classes are defined by tracks being kaons or non-kaons. The enrichment of the excited B meson states, containing an additional u, d or an s quark, is also achieved

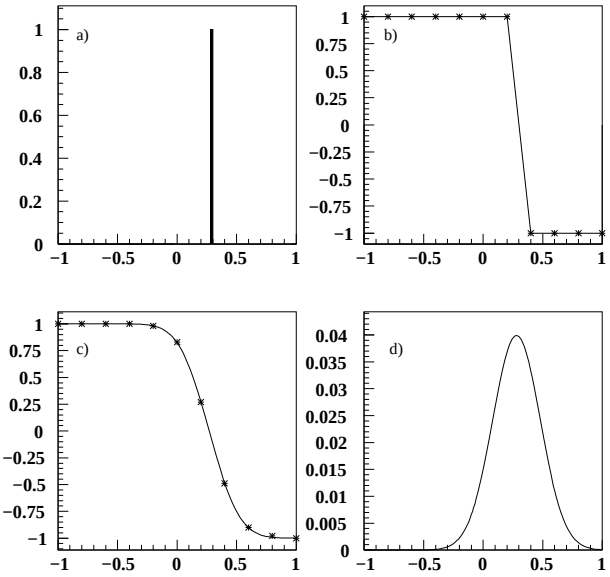

Figure 6.5: *The approximation of a probability density function for a variable lying in the range* $[-1, 1]$*. If the true variable value is 0.28 as shown in figure a), the corresponding targets of the output nodes are given by the values in figure b). Finite resolution smears out the predictions of the neural network resulting in the outputs of figure c). The outputs are fitted by a cubic B-spline to give the smooth curve. The derivative of the spline is the predicted probability for the true value to lie in the range* $[x, x + dx]$*, shown in figure d).*

by classification methods. The decay pion for the case of the $B^{**}_{u,d}$ mesons and the decay kaon for the case of the B^{**}_s meson define the signal class. Background patterns are defined by all other particles. While for the $B^{**}_{u,d}$ mesons the network approach is used for further background depletion, an analysis of the B^{**}_s meson is not possible without the use of the neural network prediction. MACRIB and the B^{**} enrichment networks are described in detail in chapter 7.

6.6.2 Probability Density Reconstruction

The advantage of the presented analysis over previous B^{**} analysis techniques is the use of a neural network method to unfold the probability density function (PDF) of the B^{**} mass distributions. The mass of a B^{**} state is dependent on the masses and momenta of its decay products, the weak decaying B and the decay pion/kaon. Both for the energy and for the direction of the B mesons (the ϕ and θ-net) predictions of their PDF exist. In the analysis chapter (8 a detailed description of these neural networks will be given.

In the following, the general idea behind the PDF prediction method is explained. The aim of PDF prediction is not just to provide a decision, as it is the case for classification problems, but to estimate a complete probability density function. With the PDF given, the parameters of interest can be easily evaluated. These are, for example, the most probable value, the mean value, the median, moments or the uncertainty intervals for providing a confidence level. To have a Bayesian interpretation the density estimation has to take into account the prior knowledge of the inclusive distribution. Thus, the prediction can never result in unphysical values, which can always happen in the frequentist approach.

Technically, the prediction of the probability density function is based on the classification decision of a number of output nodes. The target of the nodes is defined by the true parameter value and the corresponding threshold value of the output node. If the true value is above the threshold value, the target of the node is -1, and $+1$ if the true value is below the threshold value. The probability density function is the derivative of the network outputs.

As an example, consider the estimation of a quantity whose physical values lie in the range $[-1, 1]$ equidistantly distributed. The true value of a single measurement is then a random number between -1 and $+1$. If the true value is 0.28 (see figure 6.5), the target of a neural network with 9 output nodes would be $\vec{T} = (+1, +1, +1, +1, +1, +1, -1, -1, -1)$, where the thresholds of the output nodes are distributed like $\vec{\Theta} = (-0.8, -0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6, 0.8)$. If the resolution of the problem is of the order of 10%, the outputs of the nodes are smeared out $\vec{O} = (+1, +1, +1, +0.98, +0.83, +0.27, -0.49, -0.90, -0.98)$. The points $(-1, +1)$ and $(+1, -1)$ are set by default. A cubic B-spline fit through the points of the output vector provides a parametric form of the network output. In order to achieve a smooth function, the fit is regularized on the basis of the sum of the third derivatives squared. The derivative of the spline function gives the approximated probability density function for the true value. An important consideration is the number of output nodes since the best resolution the network prediction can achieve is limited by the distance of two adjacent threshold values.

6.7 Selection of Input Variables

Neural networks work by feeding in known information, which is in turn used to infer some unknown information, such as the probability of a track to be a kaon. Giving the network all information at one's disposal would result in a poor prediction performance since a great percentage of the input data would have no or little relationship with the desired output. Moreover, every additional input variable increases the dimensionality of the underlying

problem, making the minimum finding more difficult, if not impossible.

It is preferable to use only a manageably small subset of the data, including only those variables which contain significant information on the hypotheses in question. It should be mentioned that this is only valid in first order, since variables with no linear correlation to the desired output may be, in correlation to other input variables, useful.

One must also keep in mind that the training data may differ in some systematic way from the actual data. Some historical training data could not be usable since, if circumstances have changed, relationships which held in the past may no longer hold. It is also important that input variables *directly correlated* with those to be studied are not included, in order not to introduce biases.

Direct correlation means in this case that variables which could be used to analytically calculate the value to be predicted are not included. To highlight this problem an example is given. Inputs to the B^{**} enrichment networks are not allowed to have combinations of kinematic variables, which can contribute to the direct calculation of the generated Q-value. The Q-value of the B^{**} mesons is set to a specific value in the Monte Carlo simulation. If the network is able to evaluate this Q-value, an improved classification power can be reached. For Q-values out of the range of the Monte Carlo value the network predicts background and for Q-values in the vicinity of the Monte Carlo B^{**} resonance the network predicts a signal pattern. Applying the network to a flat Q-value distribution, for which one knows that no signal is included (a Monte Carlo generated background sample), a cut on background patterns would lead to a misleading signal-like peak at the position of the Monte Carlo resonance. This is a very serious bias introduced by the choice of input variables. To overcome this problem, the training of the B^{**} enrichment networks has been performed by special Monte Carlo samples, where the masses of the states are equally distributed over the considered Q-value range. The method is described in detail in section 7. The choice of variables is thus not restricted any more.

The above considerations show that for an optimal performance of the network prediction, a very careful choice of input data has to be made. The process of choosing which variables to present to the network can make up the largest part of network training and requires a deep understanding of the underlying physical process.

6.8 Efficient Network Training

In the previous sections, the basic techniques of neural networks have been described. The technique of back-propagation is conceptually simple and computationally efficient. However, there are undesirable features of back-propagation that can be avoided. In the last sections of this chapter, the

most important measures for performance improvement and for the stabilization of network training are described.

The methods described are implemented in the NeuroBayes package. The extensive use of these modern techniques throughout this analysis has considerably improved the quality of the results over analyses on the topic of excited B mesons. The improvements can not only be seen in the Q-value resolution. The time needed to set up a neural network training with stabile prediction results has been reduced by at least an order of magnitude. While network training requires hundreds of iterations to achieve a minimum prediction error, the new technique allows to reach a better minimum in only a few iterations.

6.8.1 Data Preprocessing

Rather than to represent the entire transformation from the set of input variables to the set of output variables by a single neural network function, there is often great benefit in breaking down the mapping into an initial *pre-processing* stage, followed by the parametrized neural network model itself. Raw input data can be initially transformed to a single variable. This form of pre-processing is generally called *feature extraction*. While feature extraction is mostly based on physics knowledge and is performed before starting network training by hand, there also exist automatized methods of pre-processing, which transform input variables depending on the actual distribution of the data. These automatized methods are described in the following. The distinction between the pre-processing stage and the neural network is not always clear cut, but often the pre-processing can be regarded as a fixed transformation of the variables, while the network itself contains adaptive parameters whose values are set as a part of the training process. The use of pre-processing can often greatly improve the performance of a pattern recognition system. This is a consequence of the training stage where the global minimum is searched for. This procedure can be significantly simplified if the inputs to the network are appropriately transformed. The features of network training requiring these characteristics of the data are described in detail below.

Variable Transformation

The sensitivity of the sigmoid function is, as discussed above, limited to a small input value range. Input values to the network have to be scaled accordingly into an appropriate range. Typically, raw variable values are scaled linearly. In some circumstances, non-linear scaling may be appropriate. For example, if a variable is exponentially distributed, the logarithm can be taken.

Another way of preparing the data for the neural network is to transform

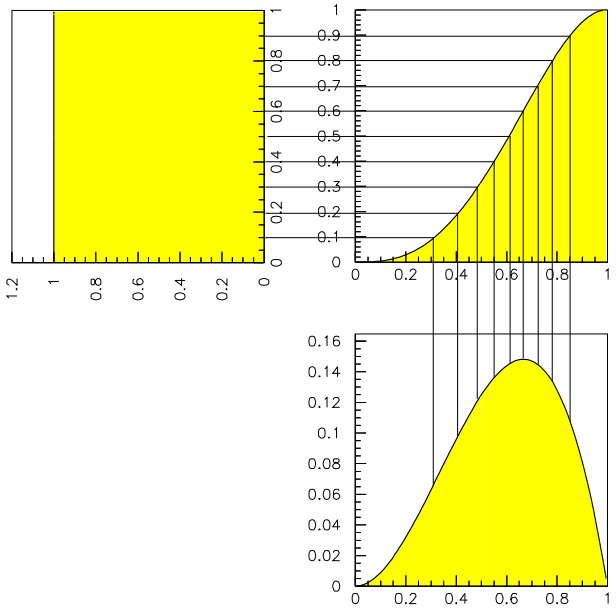

Figure 6.6: *By using the integrated function for transformation, any distribution can be equalized.*

the input variables to be equally distributed. This technique avoids gaps between measurements. If these gaps are too large compared to the size of the measurement groups, differences within values of one group disappear, i.e. they get statistically irrelevant. The transformation of the variables is performed by its integrated function. Figure 6.6 shows the procedure. If the probability density distribution of a variable is described by the function $f(x_i)$, the transformation can be achieved by the function

$$y(x_i) = \int_{-\infty}^{x_i} f(x_i')\, dx_i' \;/\; N \tag{6.30}$$

where $N =$ is a normalization factor to restrict y to the range $(0, 1)$.

To be able to give the output of a neural network a posterior probability interpretation, the input variables have to be normally distributed. The variables, which have been equalized in the previous step, can be mapped to standard Gaussians. This is performed by using the inverse of the cumulative distribution $\Phi(x)$ of the standard Gaussian. $\Phi^{-1}(x)$ can not be analytically expressed and has to be evaluated numerically. Values of $\Phi(x)$ as well as the quantiles $y(x) = \Phi^{-1}(x)$ are tabulated in many reference books (e.g. [Bra92]) and are also available from computer program libraries, e.g. the routines `FREQ` and `GAUSIN` [Lib].

Variable De-correlation

Above transformations treat single variables without taking into account any possible correlations. The variables are de-correlated by diagonalization of their covariance matrix. By dividing the new variables by the square root of the corresponding eigenvalue, the input vector space can be transformed to a hyper sphere with standard deviation of 1, centred around the origin.

If the inputs are de-correlated, the system of equations for the weights is diagonal so that it is possible to solve for the value of a single weight without concerning the other weights.

For shape reconstruction problems, the degeneracy of the new covariance matrix allows an arbitrary rotation of the basis in the input vector space. The basis is rotated so that the first variable contains all information on the first moment (the mean), of the target distribution, the second variable all information on the second moment and so on.

6.8.2 Initialization of Weights

The initial weight values are chosen to be small so that sigmoidal activation functions are not driven into the saturation regions where $g'(a)$ is very small. If the weights are too small, all the sigmoidal activation functions will be linear, which can again lead to slow training. This suggests that the summed inputs to the sigmoidal functions should be of order unity.

Suppose the input values to the network have been pre-processed so as to have zero mean and unit variance. The weights are usually generated to have also zero mean. However, the choice of variance σ^2 for the distribution of the weights is important. For a unit in the first hidden layer, the activation is given by equation 6.18. Since the choice of weight values is uncorrelated with the inputs, the average of a is zero. On the other hand, the variance of a is

$$< a^2 >= \left\langle \left(\sum_{i=0}^{d} w_i x_i \right) \left(\sum_{j=0}^{d} w_j x_j \right) \right\rangle = \sum_{i=0}^{d} < w_i^2 >< x_i^2 >= \sigma^2 d \quad (6.31)$$

where the fact has been used that the weights are uncorrelated and hence $< w_i w_j >= \delta_{ij}\sigma^2$, together with $< x_i^2 >= 1$. For a to be of order unity, the standard deviation of the distribution used to generate the initial weights should scale like $\sigma \propto d^{-1/2}$, where d is the number of connections leading to the node.

6.8.3 Weight Update Direction

The elements of the Hessian Matrix, i.e. the matrix of second derivatives of the error function with respect to the network weights, are a measure of significance of the specific weight pair. For uncorrelated input variables the

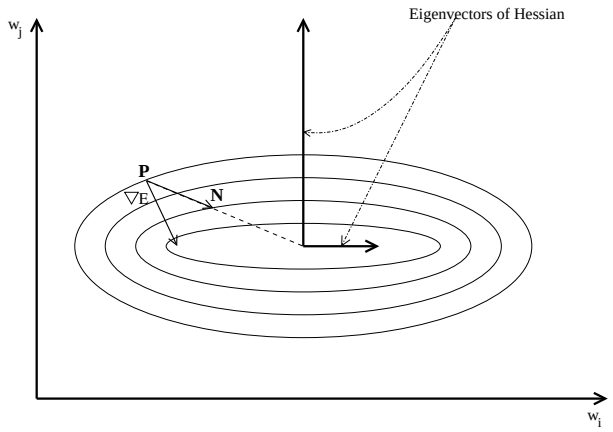

Figure 6.7: *By scaling the step size in each dimension by the inverse of the corresponding eigenvalue, the step direction can be turned towards the minimum. In general the normal gradient does not point towards the minimum as can be seen from the arrow labeled* ∇E.

matrix is diagonal, and the eigenvalues are given directly by the diagonal elements. The eigenvalues are proportional to the curvature of the error function with respect to the corresponding eigenvector. The inverse of the curvature is proportional to the significance of the given eigenvalue.

The input variables are uncorrelated by the method described above. The Hessian Matrix of the transformed variables is diagonal. The eigenvalues can be very different though. This may result in a form for the error function shown in figure 6.7, which considers the projection of the error on two weight coordinates. The eigenvectors of the Hessian Matrix point along the coordinate axes, which is a consequence of the uncorrelated input variables. The eigenvalue corresponding to the eigenvector w_i is much smaller than the one corresponding to the eigenvector w_j. The error valley is thus stretched like a canyon. For a given starting point P, the direction of weight update given by the gradient descent method is labeled by ∇E. It is obvious that this direction does not directly lead towards the minimum. Instead, if the components of the step vector ∇E are scaled by the inverse of the corresponding eigenvalue of the Hessian the resulting update direction becomes the vector denoted by **N**. A step in this direction leads directly towards the minimum.

6.8.4 Regularization

The goal of network training is not to learn an exact representation of the training data itself, but rather to build a statistical model of the process which generates the data. This is important if the network is to exhibit good

generalization, i.e. to make good predictions for new inputs. The simple analogy of curve fitting shows the effects of complexity. A polynomial with too few coefficients gives poor predictions for new data and a polynomial with too many coefficients gives poor generalization since it fits too much of the noise on the training data, the network over-learns. The number of coefficients in the polynomial controls the effective flexibility, or complexity, of the model.

A principal approach to controlling the complexity of a model is through the use of *regularization*, which involves the addition of a penalty term to the error function. The effective complexity of the model can be controlled by scaling the regularization term by an adjustable multiplicative parameter ν

$$\tilde{E} = E + \nu\Omega \tag{6.32}$$

where ν controls the extent to which the penalty term Ω influences the form of the solution. Training is performed by minimizing the total error function $\tilde{E}$, which requires that the derivatives of Ω with respect to the network weights can be computed efficiently.

The way the penalty term helps to achieve a better generalization ability can be explained by the curve fitting example. Even if the number of parameters is as high or even higher than the number of data points, the regularization term forces the result to be smooth. Ω takes the form of the sum over the second derivatives of the fitted curve at the data points. This corresponds to the integrated curvature of the function. If the multiplicative parameter ν is chosen to be too large, the fit will result in a straight line, which has zero curvature. On the other hand, if the parameter is chosen to be too low, or zero, the noise of the data is fitted and the result may show a wildly fluctuating curve. A properly adjusted value for ν can produce a fit result with a high level of generalization power, i.e. to predict function values lying between the data points used for the fit.

Similarly to the curve fitting example, one of the simplest forms of network regularizer is called *weight decay* and consists of the sum of the squares of the adaptive parameters in the network

$$\Omega = \frac{1}{2}\sum_i w_i^2. \tag{6.33}$$

The reason for calling this type of regularizer weight decay can be seen easily. For the case that the input data has no correlation to the desired output the first term in equation 6.32 is constant. The minimization of $\tilde{E}$ with respect to the weights leads to the delta-rule

$$\frac{d\mathbf{w}}{d\tau} = -\eta\nabla\tilde{E} = -\eta\nu\mathbf{w} \tag{6.34}$$

where η is the learning parameter. Here the $\Delta\mathbf{w}$ from equation 6.17 has been modified to give a differential equation in τ. This equation has the

solution

$$\mathbf{w}(\tau) = \mathbf{w}(0)\exp(-\eta\nu\tau) \tag{6.35}$$

and so all of the weights decay exponentially to zero.

On the other hand, if the data has significant information to solve the prediction problem, the weight update procedure performed by the gradient descent method lead to non-vanishing values for the weights. The interplay between the processes of weight decay and weight update per iteration leads to the optimal value for the network weights, at the point of equilibrium.

An additional form of regularization is used for the reconstruction of the probability density functions (PDF) of the B^{**} masses. One of the input components of the mass reconstruction is the neural network prediction of the B meson energy. The network prediction is applied on the basis of a previous energy measurement. This crude measurement is used to unfold the distribution of the true B energy, i.e. to assign a probability to the possible true values which could have produced the given energy measurement. The resulting function reflects the degree of belief for the value of the true parameter. The inclusive distribution of the B energy is known. There are no values below zero and above beam energy. In addition, the PDF has to be a smooth function and is not allowed to have negative values. These constraints are also included in the definition of Ω. The smoothness of the predicted PDF is given by the sum of third derivatives of the network output. The positiveness of the PDF is achieved by adding a large quadratic term to Ω if the output of a node is smaller than the output of the previous node. The constraint for the energy to be between zero and 45.6 GeV a similar quadratic term is added if the first node gives an output lower than zero and if the last node gives an output larger than the beam energy.

Chapter 7

Analysis Tools

This chapter describes the tools needed for the analyses presented in the next chapter. B^{**} mesons are contained only in b-events. Their enrichment is performed by a combined b-tag described in the first section. The analysis of the event content is performed by the software package BSAURUS [AAB+01]. BSAURUS provides a complete compilation of information essential for a variety of analyses. The outputs of BSAURUS important for this thesis are described in the second half of the chapter.

7.1 Tagging of b Events

An efficient tagging of b events is essential for the analyses presented in this thesis. Tracks originating from light quarks u, d, s and c would dominate over tracks coming from b-decays. Due to the relatively large lifetime of b-hadrons of around 1.6 ps together with the average B energy of ∼35 GeV, the mean flight distance of b-hadrons is of the order of 3 mm (figure 7.1). Compared with the DELPHI track resolution of

$$\sigma^2 = \left(\frac{61}{p \sin^{3/2}\theta}\right)^2 + (20)^2 \ \mu\text{m}, \tag{7.1}$$

the significant replacement of tracks from the primary vertex provides a unique tag for b events.

7.1.1 Primary Vertex Reconstruction

The first step of b-tagging is the reconstruction of the primary vertex. The methods of determining the de-placement of B decay tracks depend on the precise measurement of its coordinates.

The primary vertex is reconstructed for every hadronic event using the beam spot position of 200 consecutive events, which approximately corresponds to the total number of multihadronic decays in a typical Z^0 fill. The

beam spot size is rather small, ≈100 μm in the x direction and ≈10 μm in the y direction, whereas it is poorly defined along the beam direction. For light quarks, a primary vertex resolution of 22 μm and for b quarks a resolution of ≈35 μm is obtained. The somewhat degraded precision for b-events is explained by the smaller multiplicity of primary tracks.

7.1.2 The Impact Parameter Method

With the primary vertex determined, the impact parameter can be calculated. The impact parameter is defined as the minimum distance between the reconstructed track trajectory and the reconstructed primary interaction point (primary vertex). It is evaluated separately in the $r\phi$ plane and in z. The sign of the impact parameter is defined with respect to the jet direction. It is positive if the vector joining the primary vertex and the point of closest approach to the track lies in the same direction as the jet to which the given track belongs. The sign of $r\phi$ and z impact parameters is the same for a given track. The impact parameter error arises from the error on the point of closest approach and the error on the vertex position. The significance S_0 of a given track is defined as the impact parameter divided by its error.

With the given definition of the impact parameter and of its sign, it follows that the tracks from the decays of b-hadrons mostly have positive impact parameters, whereas non-zero impact parameters arising from the inaccurate reconstruction of particle trajectories are equally likely to be positive or negative. Therefore, the presence of tracks with relatively big positive impact parameters is a tag for B decay.

Figure 7.2 shows the significance for tracks of different event flavour samples. The positive side of the distribution shows a significant surplus originating from b-hadrons. This fact is used for the tagging of b events. The negative significance distribution mainly reflects the detector resolution and is used to build the track probability function $\mathcal{P}(S_0)$:

$$\mathcal{P}(S_0) = \begin{cases} \int_{S<S_0} f(S)dS, & \text{for } S_0 < 0 \\ P(-S_0), & \text{for } S_0 > 0 \end{cases} \tag{7.2}$$

which is by definition the probability for a track from the primary interaction point to have a significance with absolute value S_0 or greater. By construction, the distribution of $\mathcal{P}(S_0)$ for tracks with positive impact parameters should have a peak at low values. Using the track probability function, the probability for each track in the event can be computed according to the value of the significance. Thus, for any group of N tracks the **Lifetime Probability** is defined as:

$$\mathcal{P}_N = \Pi \cdot \sum_{j=0}^{N-1} (-\ln \Pi)^j / j! \text{ with } \Pi = \prod_{i=1}^{N} \mathcal{P}(S_i). \tag{7.3}$$

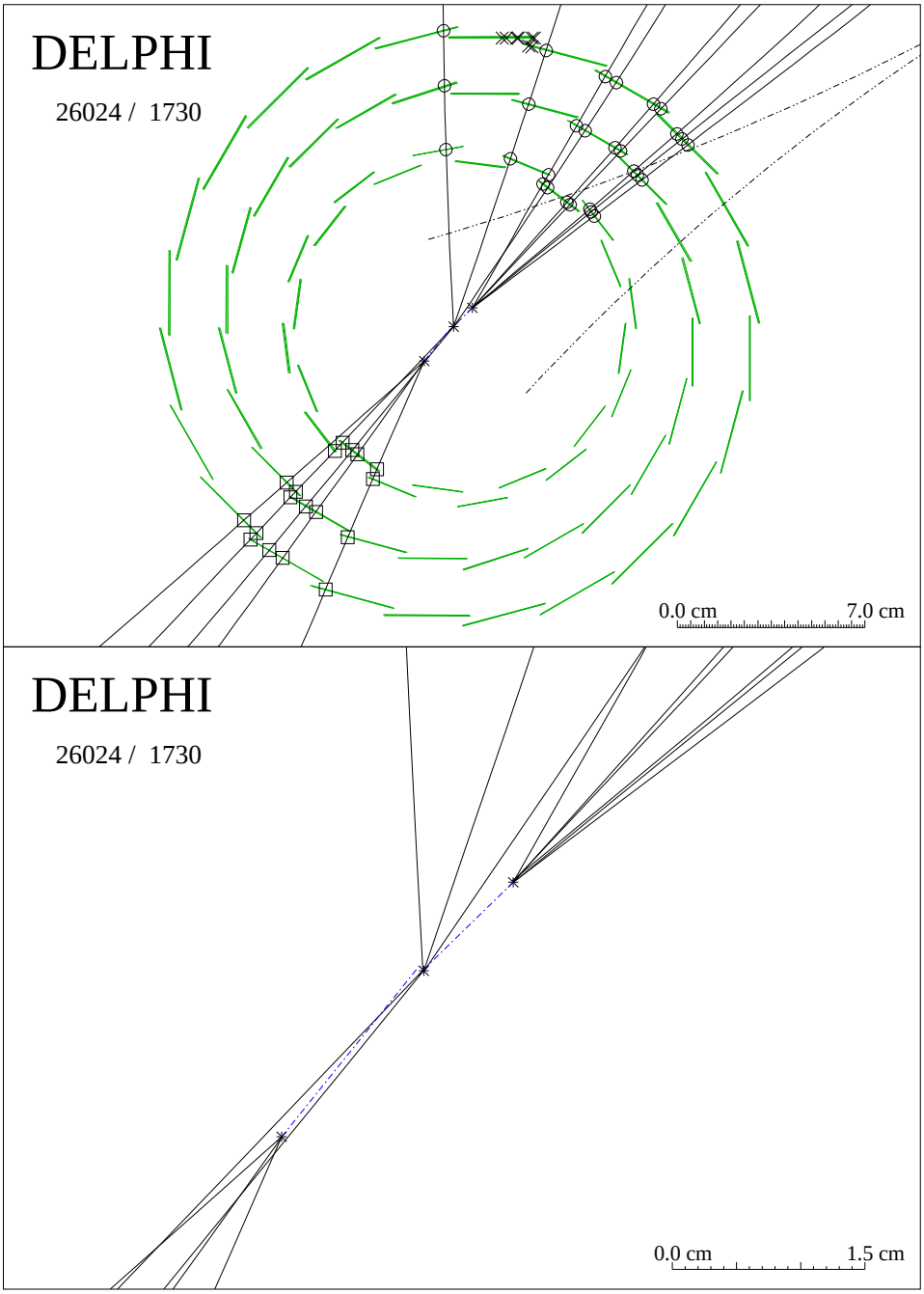

Figure 7.1: *A $Z^0 \to b\bar{b}$ Event in the DELPHI Vertex Detector. Both secondary vertices are significantly separated from the primary vertex in the middle.*

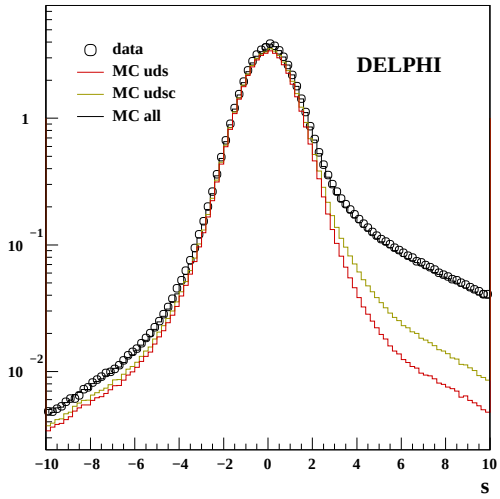

Figure 7.2: *The significance distribution of the impact parameter. The dominance of the entries contributed by b quarks on the positive side of the distribution is used for the tagging of b events.*

This variable gives the probability for a group of N tracks with the observed values of significance to come from the primary vertex. By construction, a flat distribution of $\mathcal{P}_\mathcal{N}$ is expected for a group of tracks from the primary vertex, provided the significances of these tracks are uncorrelated, while tracks from b-hadron decays should have a sharp peak at 0.

7.1.3 The Combined b-Tag

The impact parameter described in the last section is very sensitive to the b content. There is also a dependence upon other kinematic variables and the performance of the method can be increased, by using them together to form a 'combined tag'. For example, $b-$hadrons have a 10% probability of decaying to leptons and these often have a transverse momentum with respect to the b-jet axis of around 1 GeV/c or larger. On its own, the high-p_T lepton tag would have too low an efficiency but the presence of such a lepton is a useful information to combine with the impact parameter. The combined tag also makes use of other variables which have significantly different distributions for b-quark and for other events. The 'combined b-tag' is much more performant than the simple lifetime tagging, as shown in figure 7.3. The figure shows the contamination of the selected sample by other flavours versus the efficiency of b-jet selection. A very pure sample with contamination below 0.5% can be obtained for a sizeable b efficiency.

7.2 BSAURUS

Many of the tools used in the analyses presented are implemented in the DELPHI software package BSAURUS [AAB+01] developed for the inclusive reconstruction of b-hadrons in $Z^0 \to b\bar{b}$ events. The BSAURUS goal is to reconstruct b-hadron decays, by making use of as many properties of b-jets

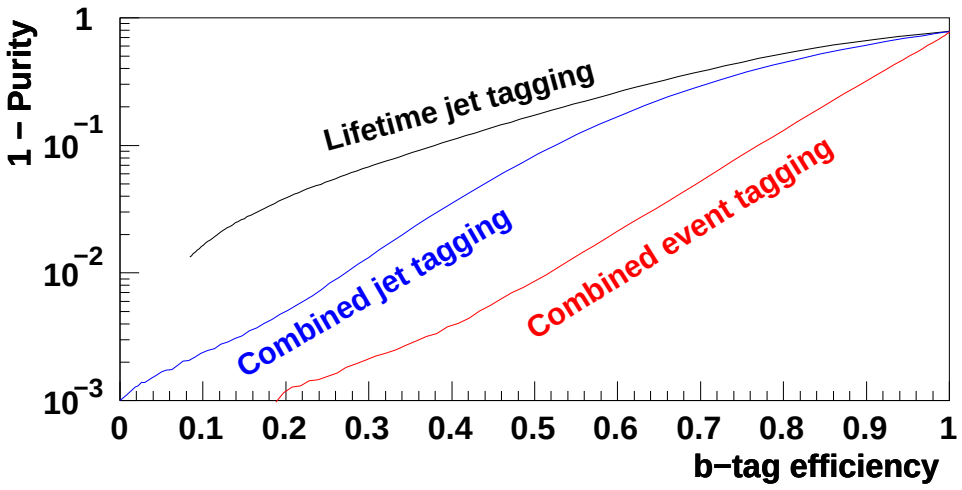

Figure 7.3: *Background suppression in Z^0 hadronic events using combined b-tagging and lifetime tagging. A considerable level of purity can be achieved for a sizeable b efficiency.*

as possible, with high efficiency and good purity. This is achieved by exploiting the capabilities of the DELPHI detector to their extreme, applying wherever possible physics knowledge about B production and decays and combining different information sources with modern tools - in particular artificial neural networks. Details of BSAURUS information important to the current analyses are described below but for a more complete treatment the reference should be consulted. The performance of BSAURUS is closely linked to the quality of event reconstruction in DELPHI and has benefited greatly in recent years from two important developments.

The first is in charged particle reconstruction. The correct assignment of silicon vertex detector points to tracks reconstructed in the TPC, the Inner Detector and the Outer Detector, has been optimized in the years 1995/1996 by better alignment procedures and the addition of completely new algorithms including e.g. inside-out tracking, a complete event ambiguity resolution, and a Kalman filter track fit taking into account multiple scattering and energy loss. All LEP I data taken from 1992-1995 have been reprocessed with these new algorithms, and this has led to huge reconstruction improvements, especially in dense jets. As an example, the signal of exclusively reconstructed D^* mesons into $D^0\pi^+$ with $D^0 \to K\pi\pi\pi$ was increased by a factor 2.4.

7.2.1 Particle Identification

The second major improvement has occurred in the area of particle identification which is crucial for B spectroscopy. The emphasis for the presented analyses lies clearly on kaon identification for the reconstruction of B_s^{**} mesons, where the main background consists of misidentified pions.

In DELPHI, hadron identification information is gathered from the Ring Imaging Cerenkov Counters (RICH) with both gas and liquid radiator, and from energy loss measurements in the TPC and the VD. Chapter 5 explains the principle measurement of the ionization effects caused by fast moving charged particles in these sub-detectors. It turns out that particle identification information from the various sub-detectors is sufficiently uncorrelated so that large gains in performance can be achieved by combining the information using a neural network technique. This approach has been implemented in the DELPHI software package MACRIB [AFM99], and is outlined briefly here.

Kaon Identification

For particles with momenta above 0.7 GeV, RICH information is combined with dE/dx measurements from the TPC. A network with a three-layer topology has been chosen, with feed-forward architecture described in chapter 6. The training included 800.000 signal patterns and the same amount of background patterns, mainly pions. The input variables provided to the network are:

- The RIBMEAN [Boua] and the HADSIGN [Boub] kaon tags, which are the alternative kaon tagging routines used by the DELPHI collaboration.
- Likelihood ratios defined by the kaon and pion probabilities provided by the liquid and gas radiators and by the dE/dx measurement of the TPC.
- Likelihood ratios defined by the expected number of photons for kaon and pion hypothesis in both radiators.
- Likelihood ratios defined by the expected Cherenkov angle for kaon and pion hypothesis in both radiators.
- Electron, muon and proton tags provided by existing packages.
- The momentum of the particle.

These inputs depend on the full functionality of the gas and liquid RICH detectors and the TPC. A separate training, without the inputs from the

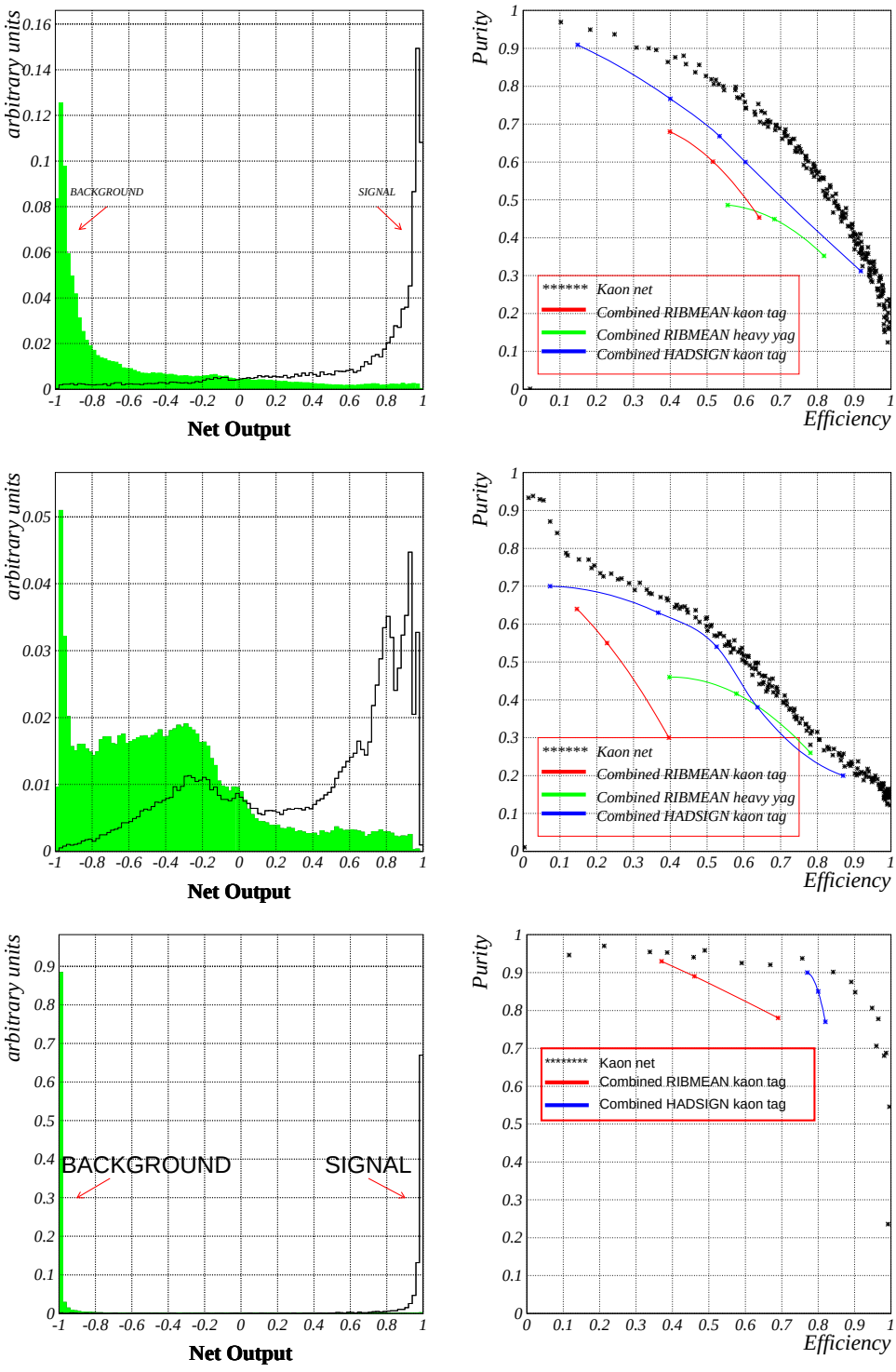

Figure 7.4: *MACRIB neural network output and efficiency/purity plots for correctly tagging a kaon for the cases where full RICH information was present (above), where liquid RICH information was missing (middle) and for the low momentum region with only VD and TPC information (below). The performance of the network approach is compared to the classical approaches of DELPHI: RIBMEAN [Boua] and HADSIGN.*

liquid radiator, has been performed for the simulation of data taken in the year 1995, a large part of which had this type of information missing.

In the low momentum region ($p < 0.7$ GeV), where RICH information is not available, dE/dx measurements from the TPC and VD are combined. They provide a very powerful separation of kaons from other particles as can be seen in figure 5.14. The training of the neural network has been performed on a sample of 5.000 kaons and the same number of non-kaons. The input variables to the network are:

- Likelihood ratios defined by the kaon and pion probabilities provided by the TPC and the VD.
- Likelihood ratios defined by the kaon and proton probabilities provided by the TPC and the VD.
- The kaon probabilities provided by the TPC and VD dE/dx measurements.
- The number of TPC wires hit by the particle.
- The momentum of the particle.

For this momentum region, no additional information from other routines on kaon, proton and lepton tagging is available. The MACRIB network approach is the first attempt in DELPHI to tag kaons with low momentum. The performance of the networks for kaon tagging for the three scenarios is shown in figure 7.4.

Proton Identification

For the protons, a very similar neural network approach has been made as for kaons [AFM99]. The performance of the networks is slightly better, though, due to an enhanced $p - \pi$ separation ability in the VD. This is due to the more distinct bands for the proton and pion hypotheses (see figure 5.14).

Electron Identification

Electron identification [KFP96] is performed using a neural network combining spatial and energy information from the HPC calorimeter, tracking information, dE/dx measurements of the TPC, and searches for kinks in the track in known layers of material and for photons radiated tangentially in them.

7.2.2 Rapidity Algorithm

The first step in the rapidity algorithm involves all reconstructed particles to calculate an event thrust axis via routine LUTHRU [Sjo]. Event hemispheres were then established by the plane perpendicular to the thrust axis and particles were clustered into event jets with the routine LUCLUS [Sjo], using a transverse momentum cutoff value of 5.0 GeV/c. A *reference axis* was then assigned to each hemisphere defined in the following way:

- If the event was a two-jet event, the reference axis for the hemisphere was the jet axis in that hemisphere.
- If a hemisphere contained 2 or more jets:
  - If one of the jets was the highest energy jet in the event, that jet axis formed the reference axis.
  - If the highest energy jet was in the opposite hemisphere, the combined probability (P_{jet}) for the tracks in a jet to have originated from the event primary vertex was formed (via the DELPHI b-tagging package algorithms [BM96, BM95]).
    The most 'b-like' jet, i.e. that with the smallest probability, was then selected to form the reference axis if $P_{jet} < 0.05$.
  - If no jet in the hemisphere satisfied the above criteria, the jet with the highest energy was selected to form the reference axis.

With the definition of the reference axis established, the rapidity of each particle was formed as follows,

$$y = \frac{1}{2} \ln \left(\frac{E + P_L}{E - P_L} \right)$$

for E the particle energy and P_L the longitudinal momentum component along the reference axis.

The rapidity variable is a powerful tool to discriminate between tracks originating from the primary event vertex and those originating from some secondary vertex displaced in space from the primary. This is illustrated in Figure 7.5 together with a comparison of the Monte Carlo with data.

By summing over all particles with rapidity greater than a cut at 1.6, an estimate of the momentum (P_y), energy (E_y) and mass ($M_y = \sqrt{E_y^2 - P_y^2}$) of the weakly decaying b-hadron is attained.

7.2.3 TrackNet

The TrackNet is the output of a neural network trained to discriminate tracks originating from the weak decay of a b-hadron from all other tracks. Important input variables to the network include the track rapidity and

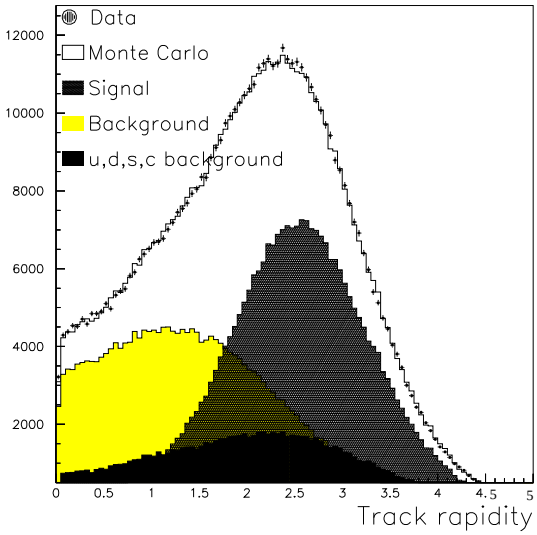

Figure 7.5: *The rapidity distribution of tracks in Monte Carlo compared to data. 'Signal' refers to tracks that originate from the B-decay chain and 'background' are tracks in b events not from the B-decay chain.*

probabilities that the track originates from the primary or a secondary vertex based on impact parameter information. The discrimination achieved is illustrated in Figure 7.6 which also shows the comparison of Monte Carlo with data. In general, secondary vertex tracks have TrackNet ~ 1 whereas tracks from the primary vertex have TrackNet values close to zero.

7.2.4 Secondary Vertex Reconstruction

The coordinates of the secondary vertex (SV) are either used to constrain the B direction reconstruction by e.g. the rapidity algorithm or, for large decay lengths of the weak decaying B meson, are directly used to estimate it by the vector pointing from the primary to the secondary vertex. Since the uncertainty in the B direction is the dominant source of error on the Q-value, this makes the precise measurement of the secondary vertex position to be the essential aspect of Q-value reconstruction of the corresponding B^{**} decay. BSAURUS provides three different methods with their own advantages and disadvantages.

The Standard BSAURUS Vertex Algorithm

In each hemisphere, an attempt is made to fit a secondary vertex to tracks with rapidity > 1.6 that pass the standard track selection criteria of section 8.1.3. From this class of track, additional criteria are applied with

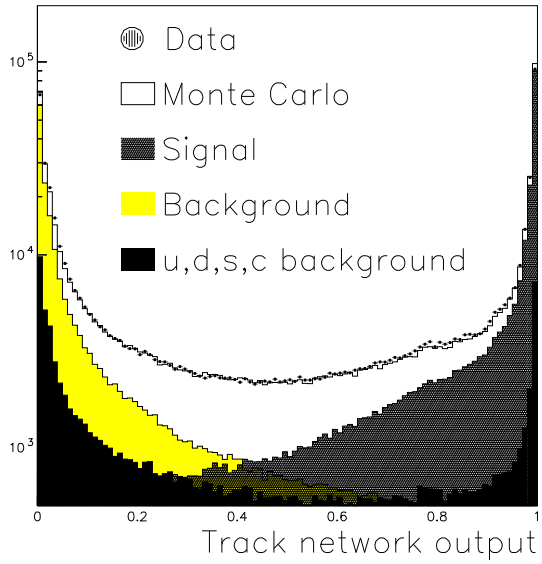

Figure 7.6: *The TrackNet distribution for tracks in Monte Carlo compared to data. 'Signal' refers to tracks that originate from the B-decay chain and 'background' are tracks in b events not from the B-decay chain.*

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. As Figure 7.7 illustrates on the left, 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. Using the selected tracks, a secondary vertex fit is performed in 3-dimensions (using routine DAPLCON from the ELEPHANT package [KFP96]) constrained to the direction of the B-candidate momentum vector. The event primary vertex is used as a starting point and if the fit did not converge[1], the track making the largest χ^2 contribution is stripped away in an iterative procedure, and the fit repeated.

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.

On the right side of figure 7.7, the decay length resolution of this algorithm is shown. Clearly visible is the rather large bias in the direction of bigger decay length. The main reason for this behaviour is the cascade D decay[2] of the B hadron leading which will 'pull' the decay length longer if

[1] Here, non-convergence means the χ^2 was above 4 standard deviations at the end of the first 10 iterations or above 3 standard deviations at the end of the next 10 iterations or took more than 20 iterations in total.

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

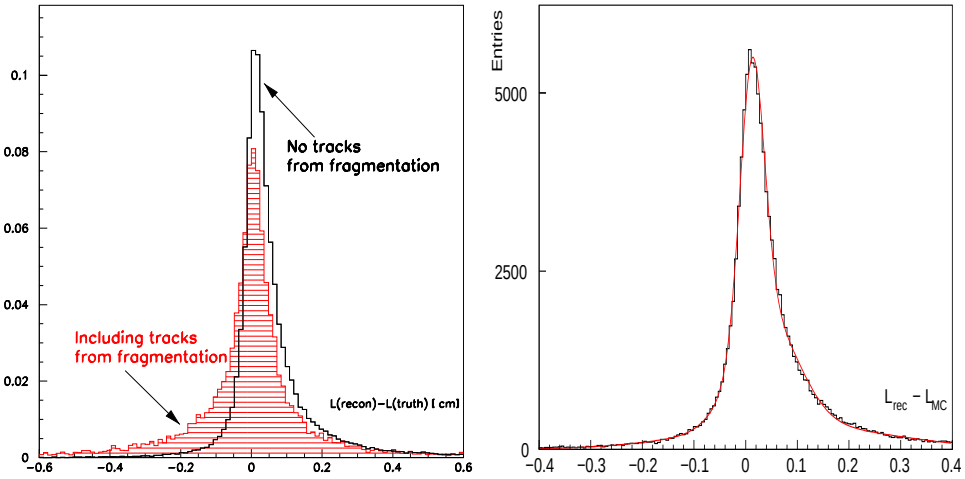

Figure 7.7: **Left:** *The pull on the decay length residual $L(recon) - L(truth)$ introduced if fragmentation tracks are allowed in the secondary vertex fit definition. The open histogram shows the case where no tracks from fragmentation are in the fit and the hatched histogram the case where fragmentation tracks are included.* **Right:** *The decay length resolution distribution for the standard vertex algorithm of BSAURUS. The rather large forward bias is clearly visible.*

tracks from the D decay are mistakenly included in the vertex fit. A clear advantage of the standard algorithm is the efficiency of almost 100% for events meeting the selection criteria of section 8.1.1.

The Strip-Down Vertex Fit

In addition to rejecting tracks coming from the primary vertex from the secondary vertex fit, an attempt has to be made to identify tracks originating from the cascade D decay. The exclusion of these tracks is performed by the BD-Net, a dedicated network for the separation of tracks originating from either the B or the D decay. The following input variables are used for the BD-Net:

- The angle between the track vector and the estimate of the B flight direction derived from the B hadron four-momentum vector.
- 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.

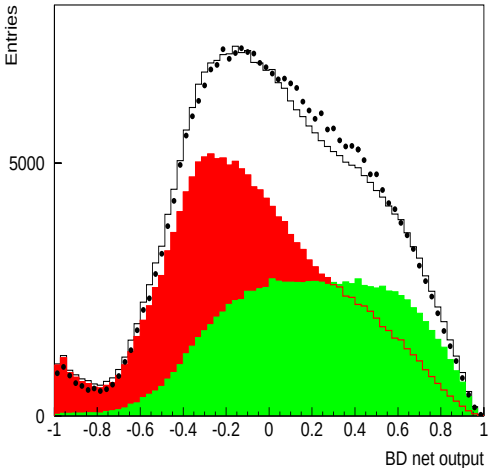

Figure 7.8: *Output of the BD-Net for tracks in simulation (histogram) and data (points) for tracks with TrackNet* > 0.5. *The component histograms show the distribution for cascade D-decay tracks (light grey) and B decay tracks (dark grey).*

- The TrackNet output.
- The MACRIB kaon identification information.
- Lepton identification information.

The output of the BD-Net is shown in figure 7.8 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 high precision.

Together, the TrackNet and the BD-Net exclude tracks from the fit that originate before and after the B vertex. Therefore, tracks are selected if they have TrackNet output bigger than 0.5 and BD-Net value less than -0.1. This BD-Net cut corresponds to 50% efficiency in selecting a track from a weakly decaying B hadron with 75% purity. Monte Carlo studies have shown that this BD-Net cut gives a much improved decay length resolution for the Strip-Down fit compared to the standard BSAURUS vertex.

A secondary vertex χ^2-fit (DAPLCON routine) is made if there are 2 or more tracks selected. If the fit fails to converge within the algorithm criteria (same criteria as in the first level fit) and more than two tracks were originally selected, the track with highest χ^2 contribution is removed[3] and the fit repeated. This procedure is done iteratively until convergence is reached or two tracks are left. The fit is constrained by the direction

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

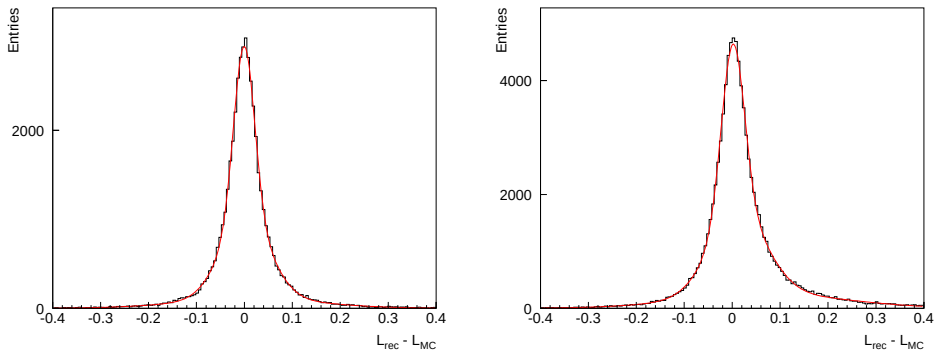

Figure 7.9: *The decay length resolution distribution for the Strip-Down vertex algorithm (left) and for the build-up algorithm (right).*

estimated from the b-hadron energy reconstruction and the starting point of the fit is the initial vertex position given by the standard algorithm.

The decay length resolution of the fit is shown on the left side of figure 7.9. One can see that the distribution is centred very nicely around 0. Also a fit of three Gaussians indicates that the resolution is nearly bias free.

Indeed, the nice performance of the algorithm does not come for free. In a lot of events, less than two candidate tracks are selected or the fit with the selected tracks did not converge. This leads to an overall efficiency for this algorithm of around 40% depending slightly on event selection criteria. Therefore, a second algorithm with higher efficiency was developed:

The Build-Up Vertex Fit

In the Build-Up method, those two tracks with TrackNet bigger than 0.5 and smallest BD-Net values are chosen to form a seed vertex. 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[4], that track with the lowest BD-Net output is added to the seed vertex definition. This is done iteratively until the remaining mass drops below the D meson mass. Finally, all the candidate tracks are fitted to a vertex.

The approach with the D meson mass is done to make sure that all tracks not belonging to the D vertex are taken into account for the b-hadron vertex fit. As a result of the fit, its decay length resolution is plotted on the right side of figure 7.9. It is clearly visible that this method leads to a bigger bias in forward direction than the stripping method. A sum of three Gaussians is fitted to the distribution. In these events, the ordering of tracks within

[4]A D mass of 1.7 GeV/c is chosen.

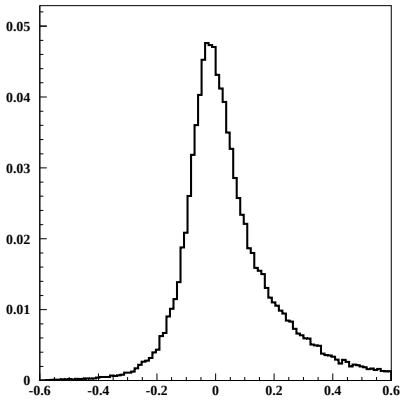

Figure 7.10: *The resolution i.e.* $(E_{rec} - E_{true})/E_{true}$ *for the weak b-hadron energy where* E_{rec} *is the reconstructed energy described in the text and* E_{true} *represents the quantity as given by the Monte Carlo generator. The distribution has a full width at half maximum of* 18.0%.

the BD-Net is not correct, leading to the incorporation of cascade D tracks in the fit and thus to a forward bias.

The advantage of the Build-Up fit method is the good overall efficiency of around 78.5% compared to approximately 40% for the Strip-Down fit.

7.2.5 b-Hadron Energy Reconstruction

Inclusive B meson reconstruction begins by dividing the events into hemispheres defined by the thrust axis. The rapidity of each reconstructed charged (assuming the pion mass) and neutral particle (assuming the photon mass) with respect to the thrust axis is calculated. The particles whose rapidities are outside a window of ± 1.6 units are considered to be B meson decay products. The momenta of these particles are added together in each hemisphere to arrive at a B meson momentum estimate for each side of the event. Studies using simulation showed that a strong correlation exists between the energy residuals $\Delta E = E_y - E_{B_{true}}$ and m_y. A fit function, characterised by 9 parameters, gives the main dependence on m_y. Finally a small *bias correction* is applied for the mean remaining energy residual as a function of the corrected energy, as determined from simulation. To reduce the amount of badly reconstructed B mesons a minimum reconstructed energy of 20 GeV in the rapidity gathering algorithm is required. This algorithm is only meaningful in two jet events and for the most energetic three jet events. Figure 7.10 shows the energy resolutions of the inclusive reconstruction method.

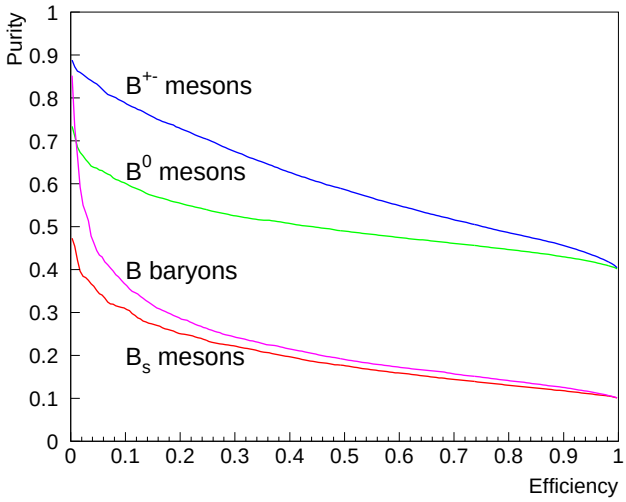

Figure 7.11: *The efficiency/purity plot of the b-hadron network. The best performance is achieved for charged B hadrons. The weakest performance is given by the B_s network for which the B_d^0 background is nearly irreducible.*

7.2.6 *b*-Hadron Four-Vectors

BSAURUS provides four different estimates of the B-candidate four-vector in a hemisphere:

- The initial estimate from the rapidity algorithm outlined in section 7.2.2.
- The three momentum vector is transformed to lie along the direction joining primary to secondary vertex (section 7.2.4). The energy component of the second estimate is the rapidity energy.
- The rapidity energy corrected 4-vector to account for neutral energy losses provides the energy component of the third estimate.
- The fourth estimate is, in principle, the same as the third but contains a more detailed energy correction 7.2.5. In addition, the momentum vector derives from the rapidity algorithm for $>$ 2-jet events that also pass $E_{\mathrm{HEM}}/E_{\mathrm{BEAM}} > 0.6$, i.e. the energy measured in the hemisphere has to exceed 60% of the beam energy to be accepted, and from the TrackNet weighted sum otherwise.

7.2.7 b-Hadron Type and Flavour

Since in a B^{**} decay the b-hadron associated to the decay track has a specific type (e.g. B^+ in the decay $B_d^{**} \to B^+\pi^-$), signal enrichment requires a sophisticated tag providing information on the hadron type. Similarly, information on the charge of the b quark contained in the weakly decaying B meson is an important discriminator for signal separation.

An optimal b-hadron type information is provided by a neural network, and its output is used as an input to the B^{**} enrichment networks, which in turn is used for signal enrichment in the analysis. While the hadron type tag is used in itself, the flavour tag is used in combination with the charge of the decay track. This combination is a very powerful input to the B^{**} networks. To account for the possibility of neutral B meson oscillation a separate flavour tag is trained for the production and decay time.

Single Hemisphere Type Identification

The construction of the variables is a chain of neural networks building up on the output of the previous one to achieve optimal performance. As a first step a neural network (SHBN) with four output nodes is trained to deliver a probability for the hypothesis that the produced meson in the hemisphere was a B^+, a B^0, a B_s or a b-baryon. The variables presented to the network are sensitive to the produced b-hadron:

- The secondary vertex charge and its error.
- The strangeness content of the hemisphere.
- The measured energy deposit of the hemisphere.
- The correlation of the charge of the leading fragmentation track with the vertex charge.

In figure 7.11 the efficiency vs. purity curves for the individual species are given. The best performance is achieved for charged b-hadrons, while B_s mesons are hard to separate from the B_d^0 background.

Production Flavour Tagging

The basis of optimal flavour 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. These probabilities are, like the type information, constructed separately for each of the b-hadron types; B^+, B^0, B_s and b-baryon.

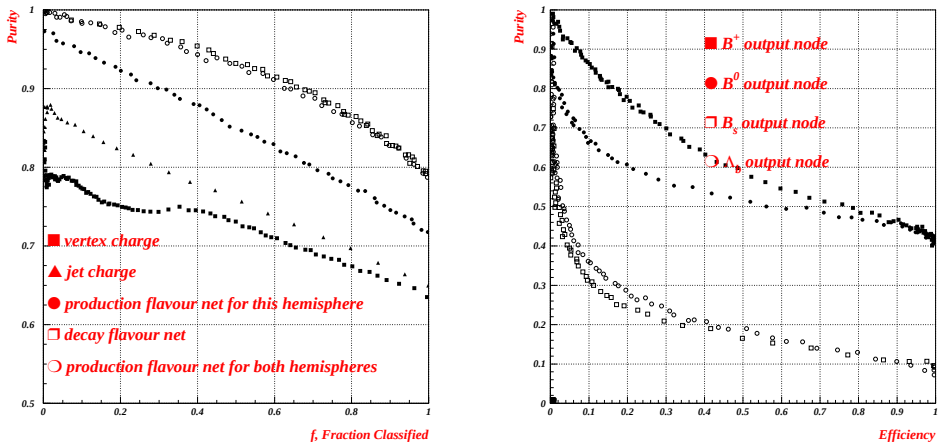

Figure 7.12: Left: *Performance of the production and decay flavour networks. The comparison is made to using only the vertex charge or jet charge variable (with $\kappa = 0.6$) as a production flavour tag.* **Right:** *Performance of the BHBN for enriching samples in the various b-hadron types. The plot is based on $b\bar{b}$ Monte Carlo making successive cuts in the network output.*

To obtain a flavour 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, are combined as the likelihood ratio

$$F(hem.)_i = \sum_{tracks} \ln\left(\frac{1 + P(same\,Q|i)}{1 - P(same\,Q|i)}\right).Q(track)$$

where $Q(track)$ is the track charge. Which tracks to sum over depends on the hypothesis considered, i.e. for the fragmentation flavour it is over all tracks with TrackNet< 0.5 whilst for the decay flavour, tracks must satisfy TrackNet≥ 0.5.

The hemisphere flavour tags form the major input to a dedicated network, the Same Hemisphere Production flavour Network or **SHPN**. This network attempts to find the optimal combination of hemisphere flavour *and* hadron type tags, for each B-species hypothesis, in order to tag the b-hadron quark flavour at **production time**. The output of the SHPN network is combined with the charge of the decay track candidate to form an important input to the B_s^{**} enrichment network and to the charge correlation network which is used to enrich B_u^{**} and B_d^{**} mesons separately. The performance of the tag can be seen on the left-hand side of figure 7.12 (full circles).

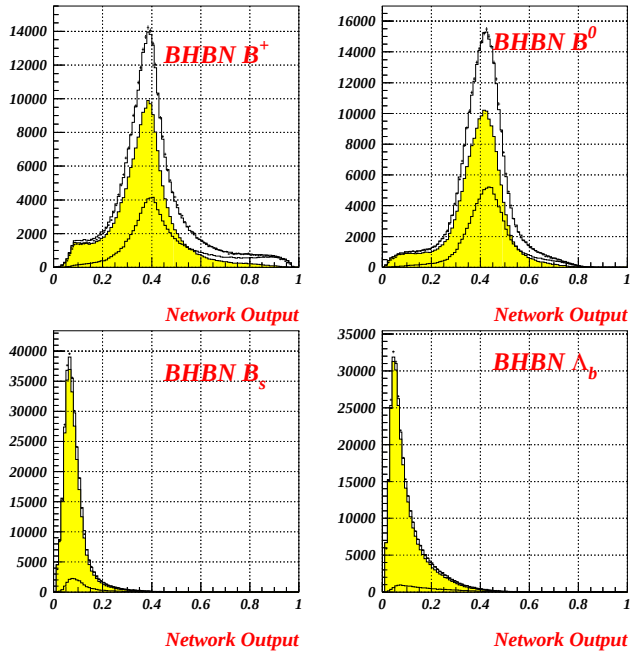

Figure 7.13: *Output of the BHBN network for the B^+, B^0, B_s and b-baryon hypotheses compared to the data (closed points). The two overlaid curves show the distribution for the hypothesis being considered i.e. the 'signal' (open histogram), compared to the distribution for everything else i.e. the 'background' (shaded histogram).*

Both Hemisphere Type Information

The performance of the SHBN can be further improved by combining the hemisphere flavour tags and single hemisphere type information introduced above. In addition, charge correlation information from the opposite hemisphere is introduced to form the Both Hemispheres B-species enrichment Network or **BHBN**. In the same way as for the SHBN, the BHBN is a network with four output target nodes, one for each B-type. The output of the network is the most important input to the B_s^{**} enrichment networks. The performance of the networks for all four hadron types is shown in figure 7.12. The output of the four nodes of the network is shown in figure 7.13. The network outputs can be interpreted as probabilities. The maximum of the B^+ and B^0 networks occurs at about 40%, which corresponds to the natural B^+/B^0 admixture in the data. In case the network output is around 40% for these mesons means that not enough information is available to tag them. Accordingly, the most probable value is taken for the probabilities, reflecting the average probability.

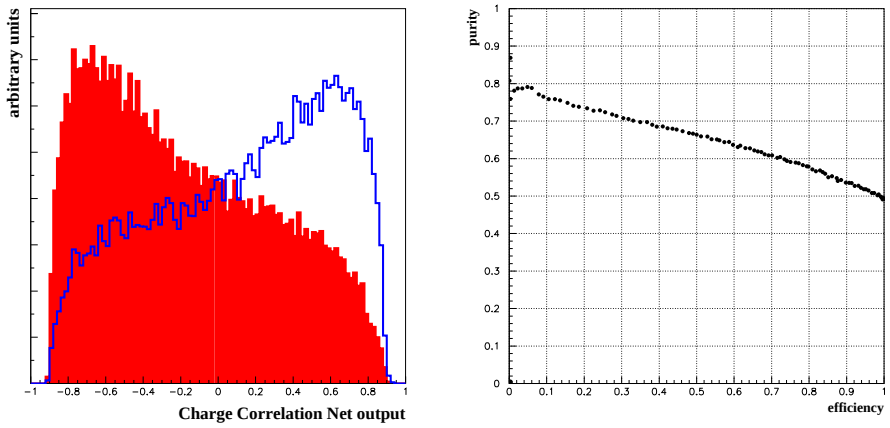

Figure 7.14: *The performance of the Charge Correlation Net. (a) The network output for negative correlation (solid line) and for positive correlation (filled area). (b) The purity vs. efficiency of the Charge Correlation Net.*

Decay Flavour Tagging

The Both Hemispheres Decay flavour Network or **BHDN** utilises similar constructions for the discriminating variables as in the SHPN, the main difference being that opposite side hemisphere information is explicitly used in order to boost the performance, which is also shown in figure 7.12 (open squares) on the left-hand side.

7.3 The Charge Correlation Net

The measurement of the isospin of the B^{**} states requires the ability to determine the correlation of the track charge with the flavour of the weak decaying B state. The Charge Correlation Net allows to divide tracks entering the Q-value distribution in two subsamples with the right and wrong correlation to the B. A separate determination of the B^{**} rates in both subsamples can supply a confirmation of the $B^{**}_{u,d}$ and B^{**}_{s} isospin. The Σ_b^+ and the Σ_b^- states are not separable so that the classification of tracks into the correlation samples can deliver only the equality of the rates.

The network was trained on the correlation of the track charge with the b-quark flavour. The network is constructed with a topology of 20 nodes in the hidden layer and one output node. The input variables of the net, multiplied by the track charge, are:

- The single hemisphere B^+ decay flavour net (DFLBP) combined with the single hemisphere production flavour net of the other hemisphere.

This has the advantage of excluding the B^{**} track candidate from the determination of the B^+ decay flavour, since the DFLBP net does not include fragmentation tracks.

- The jet charge minus the jet charge of the opposite hemisphere.
- The weighted number of pions by their charge and the probability to come from behind or in front of the secondary vertex. For example, for the $B^+ \to D^- \pi^+ \pi^+$ decay one expects positive pions behind the secondary vertex and negative in front.
- The both hemisphere decay flavour net.
- The likelihood-ratio-based production flavour tag for B_s and B^0.

Chapter 8

The Analysis

This section presents the search for and analysis of orbitally excited B mesons containing u, d and s quarks. The analyses rely on the precise reconstruction of the decay Q-value achieved by a neural network technique. The enrichment of the states is performed by dedicated B^{**} networks exploiting the special kinematics of B^{**} decay tracks and the correlation between the flavour of the weakly decaying B meson and the charge of the primary particle. The B^{**} networks provide the probability of the track (π/K) to originate from the B^{**} decay. The B_u^{**}, B_d^{**} and B_s^{**} states are treated separately allowing the possibility of the independent measurement of the corresponding rates.

8.1 Data Sample

For the analyses, data taken at the Z^0 peak in 1994 and 1995 have been used. In addition to the LEP I data, the detector calibration runs on the Z^0 peak in the years 1996 to 1998 have also been considered. The analysis starts with a multihadronic event selection, followed by the tagging of b-events, with the additional requirement that the b-jets lie within the barrel part of the detector. In addition, the used tracks have to pass a set of standard quality cuts listed below.

8.1.1 Selecting Multihadronic Events

Multihadronic Z^0 decays were selected by the following requirements:

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

These requirements resulted in an efficiency estimated to be 92.6% and all backgrounds to be below the 0.1% level in the data.

8.1.2 Event Hemisphere Selection

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

- $|\cos\Theta_{thrust}| < 0.7$
- The secondary vertex fit converged successfully according to limits set in BSAURUS on the fit χ^2 value and the number of minimization iterations needed.
- $0.5 < (x_{hem} = E_{hem}/E_{beam}) < 1.1$

In addition, the data are enriched in $Z \to b\bar{b}$ events by using the standard DELPHI b-tagging package [BM96] based on track impact parameter information. The tagging variable from Monte Carlo is tuned to the data by a smearing procedure applied to track impact parameters [BM96]. These adjusted track parameters are also used by default in the BSAURUS package. The event b-tag was cut at 0.0 providing a minimum b-purity of 90% for all analyses. Strong enrichment of B^{**} mesons lead naturally to a better b-purity too.

8.1.3 Track Analysis Cuts

All tracks used in the BSAURUS package and subsequently in the separate analyses are required to pass the following set of standard track quality cuts:

- Impact parameter in the $r\phi$ plane $|\delta_{r-\phi}| < 4.0$ cm
- Impact parameter in the z plane $|\delta_z| < 6.0$ cm
- $|\cos\theta| < 0.94$
- $\frac{\Delta E}{E} < 1.0$
- At least one $r-\phi$ track hit from the silicon vertex detector (VD)
- Tracks must not have been flagged as originating from interactions with detector material by the standard DELPHI interaction vertex reconstruction package described in [A$^+$00a].

The number of events remaining after the subsequent selection phases is presented in table 8.1. Also included is the number of simulated events after the corresponding selection phases. Each event can be subdivided into two hemispheres corresponding to the b and the $\bar{b}$ quark. The BSAURUS

Selection	Data			Simulation	
	1994	1995	1996 - 1998	1994 $q\bar{q}$	1994 $b\bar{b}$
Multihadron	1400250	677507	130392	2993119	1753723
$\cos(\theta_{thrust})$	837986	402540	77262	1798506	1050973
b tagging	166103	79591	14855	353947	851338
Hemispheres with converged SV fit	276542	132716	25008	601435	1476791

Table 8.1: *The number of events selected after various cuts. The last row shows the number of hemispheres with a converged secondary vertex fit in the remaining events. Note that this number is approximately double the number of events in the second last row, except for efficiency losses due to the vertex fit.*

secondary vertex algorithm described in the last section attempts to fit the decay vertex of the weak B. The hemispheres where this fit has converged are taken for the analyses. The number of attained hemispheres in the different years of data taking and also in the Monte Carlo samples are summarized in the table.

8.2 Monte Carlo Samples

The Monte Carlo sample of $Z \to q\bar{q}$ events, details of which are listed in Table 8.3, was approximately 3 times the statistics of the data. The generated events were passed through a full detector simulation [DEL95] and the same multihadronic selection criteria as the data. Weights were applied to the simulation to correct for the following effects:

- The current world average measurements of B_s and Λ_b lifetimes and b-hadron production fractions as listed in Table 8.2.
- The b-fragmentation function. The value of $\langle x \rangle$ in the default Monte Carlo ($\langle x \rangle = 0.708$) was weighted to agree with the value obtained from a recent DELPHI analysis of the 1994 data set ($\langle x \rangle = 0.7153 \pm 0.0052$ [GMaL02]).
- Hemisphere quality. A hemisphere 'quality flag' which was correlated to the number of tracks in the hemisphere likely to be badly reconstructed e.g. those tracks failing the standard selection criteria of section 8.1.1. The weight was constructed to account for data/simulation discrepancies in this variable and was formed in bins of the number of charged particles in the hemisphere, that passed the standard quality cuts of Section 8.1.1, thus ensuring that the overall good charged particle multiplicity was unchanged after the application of the weight.

b-hadron species	Lifetime	Production fractions $f(b \to B_x)$		
		Values	$\rho(B_x, B_s)$	$\rho(B_x, \text{baryon})$
B_s	1.464 ± 0.057ps	$f(\text{B}_s^0) = 0.097 \pm 0.011$	-	+0.034
b-baryons	1.208 ± 0.051ps	$f(\text{baryon}) = 0.104 \pm 0.017$	+0.034	-
B^0 or B^+	-	$f(\text{B}^0) = f(\text{B}^+) = 0.399 \pm 0.010$	-0.577	-0.836

Table 8.2: *Values for the b-hadron lifetimes and production fractions (together with correlations) used to re-weight the Monte Carlo [OG01].*

8.2.1 Special Monte Carlo Production

In addition to the standard DELPHI Monte Carlo sample (MC), a special sample was created for the purpose of neural network training. For the simulation, the same generator has been used (JETSET 7.3) as for the standard MC files. These files are special in the way that only signal events have been generated with the Q-value of the B^{**} decays smeared over a large range compared to the expected width of the real states. Using these events for training prevents the network from learning the explicit position of the simulated distribution. The performance of the networks, though, has been in each case evaluated on the statistically independent sample of normal simulated events, before applying it to data and carrying out the analyses described below.

8.3 The Experimental Method

The reconstruction of the B^{**} decay Q-value and the enrichment of signal tracks in the Q-value distributions are the key elements of the analysis. Both procedures are attained via a neural network approach providing the best possible performance. The idea behind the approaches is outlined below. The inputs to the specific networks corresponding to the B_u^{**}, B_d^{**} and B_s^{**} analyses are described in the appropriate sections.

8.3.1 Q-Value Reconstruction

Candidate tracks entering the analysis procedure are combined with the reconstructed B meson of the hemisphere and the Q-value of the decay is evaluated. The easiest way of reconstructing the Q-value is to combine the four-vector of the decay pion with the four-vector of the weak decaying B meson provided by BSAURUS. This leads, however, to two distinct problems: firstly, the secondary vertex and thus the B direction are estimated by a weighted sum over tracks, taking into account their probability to belong to the secondary vertex, given by the TrackNet output of BSAURUS. While this ansatz provides a good overall performance for all hemispheres, it is not optimal for the spectroscopy of B mesons. The reason for this is easy to see: decay tracks originating from B^{**} decays have a relatively high rapidity, like B decay tracks, and can be thus contained in the reconstruction of the

Monte Carlo Generator	JETSET 7.3 [Sjo]
Quark Masses	
m_t	50.000 GeV
m_b	5.000 GeV
m_c	1.600 GeV
m_s	0.500 GeV
m_u	0.300 GeV
m_d	0.300 GeV
Resonance Parameters	
Z^0 mass	91.187 GeV
Z^0 width	2.489 GeV
Perturbative ansatz	Parton shower
Λ_{QCD}	0.346 GeV
Q_0	2.25 GeV
Non-perturbative ansatz	String fragmentation
σ_{p_T}	0.428 GeV
γ_s	0.28
Diquark rate	0.1
Heavy quark fragmentation function	Peterson et. al. [PSSZ83]
ϵ_b	0.002326
ϵ_c	0.030476
Light quark fragmentation function	Symmetric LUND [And97]
a_α	0.354
a_β	0.523
b-hadron production rates	
B_u	39.9 %
B_d	39.9 %
B_s	11.1 %
b-baryon	9.1 %

Table 8.3: *Details of the Monte Carlo generator used together with some of the more relevant parameter values that have been tuned to the DELPHI data [A+96a].*

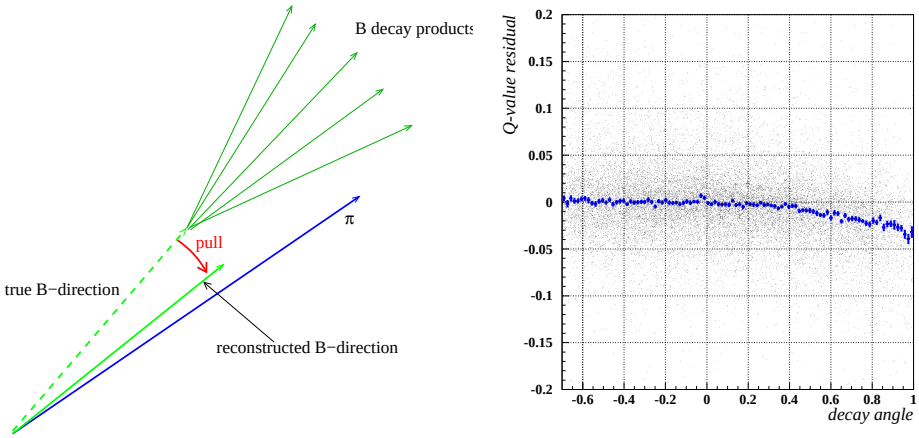

Figure 8.1: Left: *A schematic view of a B^{**} decay into a pion and a b-hadron. The presence of the signal pion effects the reconstruction of the B-direction. This effect is enhanced if the decay angle in the B^{**} rest frame is very forward with respect to the B^{**} flight direction, as can be seen in the plot on the* **right**. *The distribution shows the resulting Q-value residual as a function of the π decay angle.*

B four-vector. This results in a severe **bias** in the Q-value distribution of signal tracks. The effect of this bias is illustrated in figure 8.1.

The second problem that arises is that a measurement error cannot be associated to the Q-value measurement. The corresponding error in the framework of linear deviations has been derived in equation 4.4. There it has been mentioned that normal error propagation is not appropriate in the case of spectroscopy. The obvious kinematical reason for this is shown in figure 8.2. For a given error on the direction of the B meson, the opening angle α between the measured B and the π momentum vector can vary considerably, if, and this happens often, the pion momentum lies in the error cone of the B. This effect leads to a highly non-linear nature of the Q-value error requiring a non-linear approach for solving.

This non-linear approach is provided by a neural network ansatz, that provides a correction on an existing Q-value estimate. The starting point is the value Q_0 obtained by:

- The B direction estimate given by the primary to secondary vertex vector. The secondary vertex coordinate is provided by BSAURUS in form of the strip-down vertex. If the strip-down vertex is not defined, the build-up vertex is used, and if the build-up vertex is not defined, the standard BSAURUS vertex is taken, which is always defined once the selection criteria of sections 8.1.1 and 8.1.2 are met.

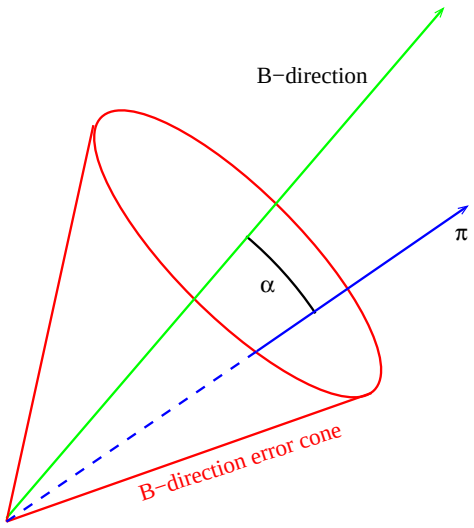

Figure 8.2: *The kinematic reason for the non-linearity of the Q-value error. For the case that the π is inside the error cone of the B-direction estimate, the Q-value can change drastically for small deviations of the opening angle α.*

- The B energy estimate given by the sum of particle energies with rapidity above 1.6.
- the 4-vector of the candidate track.

The direction given by the strip-down or build-up algorithms is optimal because for the definition of the vertices only tracks with TrackNet **above** 0.5 have been chosen. At the same time candidate tracks for B^{**} decay are only considered if their TrackNet output is **below** 0.5. Since the candidate tracks can by definition not be included in the vertex, a bias free direction estimate is possible. An additional advantage of using the primary to secondary vertex direction is the better B-direction resolution compared to the standard BSAURUS B-vectors. The B energy estimate completing the B 4-vector is found to introduce a small bias in the Q-value due to the inclusion of the signal track via the rapidity weighted sum, but is, at the same time, the best compromise with regard to the energy resolution.

The Q-Value Correction Network

The original Q-value estimate can be much optimized by a correction supplied by the Q-Value Correction Network or **QCorr** on a track-by-track basis[1]. This network approach reduces the effect of biases arising in the

[1] At the same time as the Q-value resolution gets better for signal tracks from B^{**} decays, the resolution of background tracks gets worse. This is not a problem as long as

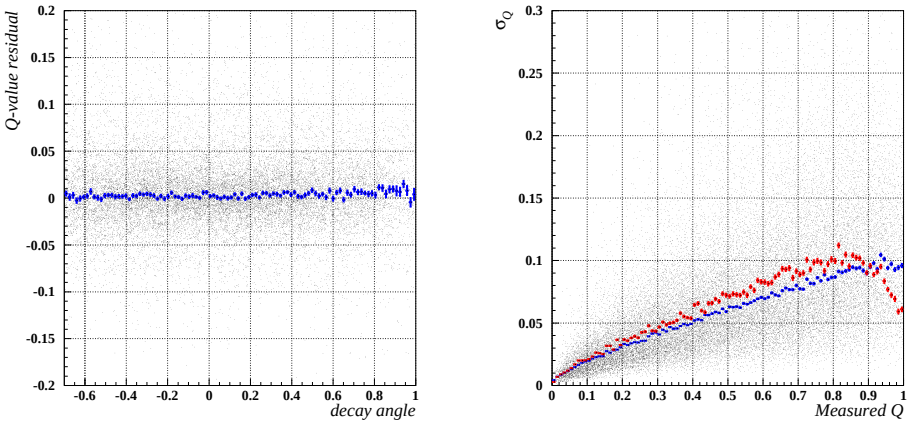

Figure 8.3: Left: *The distribution shows the Q-value residual as a function of the π decay angle after the application of the QCorr network. The bias of figure 8.1 is not present any more.* **Right:** *σ_Q as a function of the Q-value. The blue (dark) points represent the median of the distribution. Superimposed is the deviation from the true Q value in red (light), showing a good agreement with the estimate. The break-down at high Q-values is due to effects caused by the nature of conditional probability densities.*

energy input or even in any combination of inputs and at the same time improves the Q-value resolution. The corrected Q-value residual can be seen on the left-hand side of figure 8.3.

The output provided by the network is not just a specific value but a full **probability density function** (or pdf). The technical details of reconstructing pdfs of variable distributions via neural network techniques is described in detail in section 6.6.2. A schematic picture of the network output is shown in figure 8.4. The original Q-value Q_0 is located at the centre. The pdf of the network output indicates not only the most probable value of Q_{true} but gives due to its width also a measure for the reliability of the estimate. The application of the network has shown that the width of the pdf is the most powerful result of the QCorr network providing a track-by-track estimate of the measurement resolution. Compared to this, the absolute value of the resolution improvement has proven to be small, since the corrections on Q_0 are also small.

The pdf shown in figure 8.4 is non-Gaussian and has a long tail towards large positive values. The best position estimator as well as the best width estimator of general pdfs is not defined and has to be adjusted to the spe-

the resulting background distribution in the corrected Q-value is not peaked, thus faking a resonant structure. This has proven to be not the case.

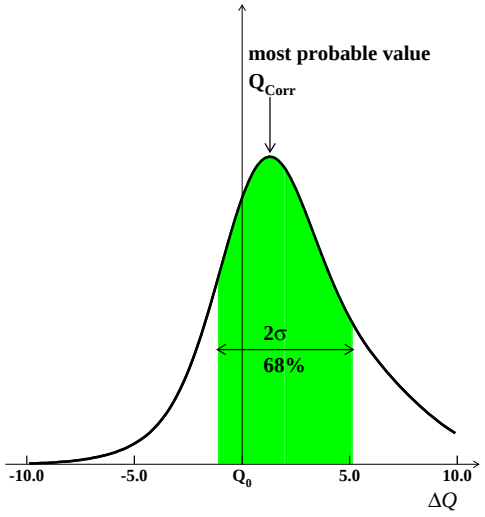

Figure 8.4: *A schematic picture of a possible output of the QCorr Network.*

cial case under consideration. For the Q-value correction ΔQ the mode of the probability density has been chosen, given by the maximum likelihood estimator. For the width, half of the 68% confidence interval around the median is taken, and in analogy with Gaussian distributions called σ_Q.

The Network Definition

The training of the network has been performed on the Monte-Carlo sample of section 8.2.1. The flatly distributed Q-value of the B^{**} mesons preserves the network from 'learning' the true Q-value and thus introducing a bias. The networks have been trained on the correction factor

$$\Delta Q = \frac{(Q_{rec} - Q_{true})}{\sigma_{Q_{rec}}} \tag{8.1}$$

where Q_{rec} is the original Q-value estimate to be corrected on, Q_{true} is the generated Monte-Carlo quantity and $\sigma_{Q_{rec}}$ is the error estimate on the measurement. The error estimate is based on Monte Carlo studies of the Q-value residual as a function of the reconstructed Q-value. The functional dependence of σ_Q on Q can be described as $\sigma_Q \propto c\sqrt{Q}$ with c adjusted to the specific analysis ($B^{**}_{u,d} or B^{**}_s$). The network provides therefore a correction on a track-by-track basis. Separate networks have been trained for the reconstruction of $B^{**}_{u,d}$ and B^{**}_s mesons thus taking into account the mass ranges of the excited states and their different decay channels. B^{**}_s mesons were generated with masses in the range 0 to 400 MeV Q-value (for the decay

into BK) and $B^{**}_{u,d}$ mesons with masses in the range 0 to 1 GeV (for the decay into $B\pi$). In the following, track based input variables are evaluated using pion/kaon hypothesis accordingly. The input variables used to train the networks are:

- The original Q-value.
- The corresponding energy of the weak B meson and, in addition, the b-hadron energy estimates provided by following algorithms: the rapidity algorithm, the TrackNet weighted and the most b-like jet.
- The track decay angle in the B^{**} rest frame with respect to the B^{**} flight direction.
- The TrackNet output of the track.
- The gap in the TrackNet output between the two particles on either side of the TrackNet = 0.5 point. This quantity is a measure of separation between the primary and secondary vertex, in terms of the TrackNet algorithm.
- The rapidity of the track.
- The opening angle α between the b-hadron and the track candidate as measured in the lab frame.
- The Q-value estimates evaluated with all four B four-vectors provided by BSAURUS (see 7.2.6).
- The same quantities with the four-vector of the track subtracted. Due to the weighting procedure to evaluate the level of affiliation of the track to the weak B, the signal candidate is never completely in or out of the vertex. The answer is somewhere in between. The last two input quantities, thus, give the range in which the true Q-value is possibly lying.
- The decay length, its error and the significance obtained from the best existing vertex.
- A variable informing the network which BSAURUS algorithm was used for the secondary vertexing. This is necessary since the various algorithms do not have the same resolution.
- The error on the ϕ and θ component of the B direction vector resulting from the standard secondary vertex fit of BSAURUS.
- Weighted sums of charged and neutral particles to estimate the charged and neutral component of the b-hadron energy.

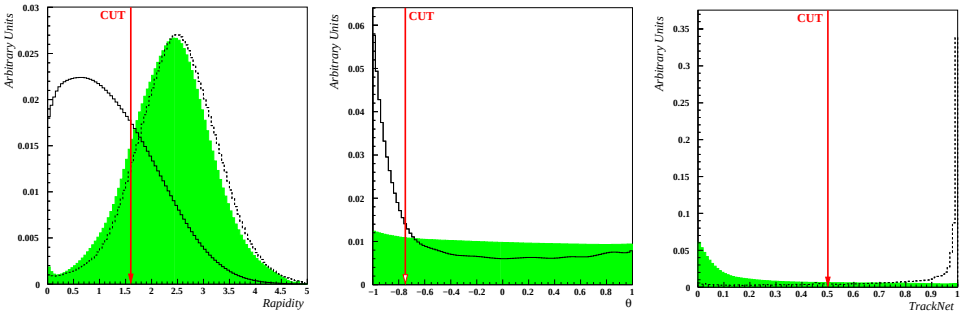

Figure 8.5: *The distributions cut on prior to network training. On the left, the rapidity distribution of fragmentation tracks (solid line) is compared to the distribution of B-decay tracks (dotted line) and signal tracks (green,shaded). In the middle, the decay angle distribution in the B^{**} rest frame is shown for signal tracks (green,shaded) and for fragmentation tracks (solid line). On the right, the TrackNet output for signal tracks (green,shaded) and for B-decay tracks (dotted line) are plotted.*

- The secondary vertex mass.
- The rapidity mass.
- The hemisphere quality.
- The event thrust value.
- The total energy reconstructed in the hemisphere.
- The information if a high p_T lepton has been present in the hemisphere.
- The estimate of missing p_T between the B-direction and the thrust axis calculated using fragmentation tracks from the same hemisphere and all tracks from the opposite one.

Network Precuts

The precuts made prior to network training and at network application are listed below and illustrated in figure 8.5. Additional cuts have only been made for the individual B^{**} analyses and are stated in the particular sections later on. The importance of precuts prior to network training has been stressed in section 6.7. The cuts applied were:

- Due to the strong decay of the B^{**} mesons, the primary pions originate from the primary vertex and not from the secondary vertex. At the same time, the primary pions have large rapidities, just like B decay tracks. Fragmentation tracks can be depleted by applying a rapidity

cut at 1.6, motivated by the BSAURUS track selection criteria for tracks originating from the secondary vertex.

- This cut depletes most of the fragmentation background but has little effect on B decay tracks. To reduce this source of background, the information provided by the TrackNet routine is used. Tracks with a TrackNet output above 0.5 are considered to originate from the secondary vertex and are not further analyzed.

- Since the B^{**} states are assumed to be unpolarized, their decay is expected to be isotropic. Fragmentation tracks have low energies compared to the B^{**} and thus appear to be backward with respect to their flight direction. The decay angle in the rest frame of the B^{**} shows a flat distribution for signal tracks whereas it is strongly peaked at $\cos\theta^* = -1.0$ for fragmentation tracks. To further deplete fragmentation tracks, a cut at $\theta^* > -0.75$ has been applied.

8.3.2 B^{**} Enrichment

Accounting for the production ratios given in the simulation of B_s^{**} mesons, the track-level signal purity $P \approx 0.15\%$ averaged over the whole Q-value region[2]. The evaluation of this quantity takes into account the production ratio of $q\bar{q}$ pairs, their tagging efficiencies, the efficiency of multihadron selection, the rate of B_s^{**} production and the mean multiplicities in the corresponding multihadronic decays.

This B_s^{**} signal purity was the major motivation for the development and usage of neural network techniques for signal enrichment for all B^{**} analyses and not exclusively for B_s^{**} mesons. The method is based on:

- The exploitation of the special kinematics of B^{**} decay tracks.

- The correlation between the flavour of the weakly decaying B meson and the charge of the primary particle, provided by the Charge Correlation Network (or **ChCorr**)

- The excellent enrichment power of the b-Hadron Type Identification Networks (or **HadId**).

- Particle identification provided by the MACRIB package.

The combination of these provides a track-by-track measure that the track originates from a B^{**} decay. The main goal is firstly to separate tracks coming from the primary vertex from those coming from the secondary vertex and secondly to separate signal tracks from fragmentation tracks.

[2]The given purity corresponds to the purity of signal kaons originating from B_s^{**} decays without any cuts applied on the $q\bar{q}$ Monte Carlo simulation.

The ansatz followed for the enrichment of $B_{u,d}^{**}$ mesons on one side and for the enrichment of B_s^{**} mesons on the other side differs in details. $B_{u,d}^{**}$ are enriched in several steps. First the common kinematical information of B_u^{**} and B_d^{**} mesons are exploited to enrich both in a mixture, and second the charge correlation and b-type information is used to separate them. The enrichment technique of B_s^{**} mesons incorporates the same components with the difference that all information is fed into one single network. The MACRIB kaon net plays a crucial role in the B_s^{**} enrichment network since the decay kaons are almost fully masked by the vast number of pions produced in an event. The specific enrichment techniques, together with the network input variables and the precuts, are discussed in the corresponding sections of the different B^{**} analyses.

8.4 A Study of the B_d^{**} and B_u^{**}

If the mass of the B^{**} mesons is above the $B^*\pi$ but below the $B\rho$ threshold, the B^{**} mesons decay mainly into $B\pi$ or $B^*\pi$. Following the predictions of HQET, B^{**} mesons are expected to be grouped in two doublets, one narrow and one broad. The decays of the states occur into B or to B^* according to their quantum numbers. The broad B^{**} states have never been directly observed experimentally. Any hint in the data towards these states would lead to a major experimental progress. The decay of the narrow states is considered in the charged mode $B^{**} \to B^{(*)+}\pi^-$ only because of the very low π^0 efficiency achieved by the DELPHI detector.

Note that with the present experimental methods there is no distinction between B and B^* mesons. Therefore, they are labelled as $B^{(*)}$. Whether the decay was actually into a B or a B^* has a negligible effect on the resolution in the Q-value. However, the decay of a B^{**} of a given mass to a $B^*\pi$ gives rise to a Q-value which is shifted downwards by the $B^* - B$ mass difference of about 46 MeV/c^2 compared to the Q-value of a $B\pi$ decay. Therefore, for the determination of a B^{**} mass from the measurable Q-value, an assumption about the B^*/B ratio in the signal has to be employed.

8.4.1 The Q-Value

The Q-Value Correction Network, optimized for $B_{u,d}^{**}$ decays, provides a Q-value estimate for each candidate together with the corresponding resolution. The estimated resolution is dominated by energy and angular resolution of the B meson. The training of the $B_{u,d}^{**}$ QCorr network has been performed on a special Monte Carlo sample, which has a flat Q-value distribution for B^{**} decays. The Q-value dependent measurement error σ_Q is shown in figure 8.6 with the median of the distribution superimposed.

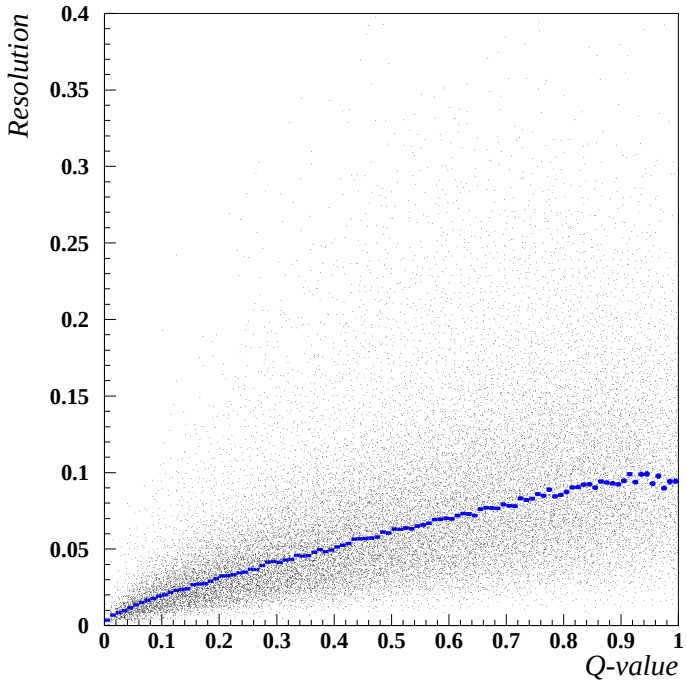

Figure 8.6: *Median of the Q-value resolution σ_Q as a function of the Q-value. At the anticipated signal region at approximately 280-290 MeV, the median resolution is expected to be $\sigma_Q \approx 34$ MeV (averaged over B_u^{**} and B_d^{**} mesons which have slightly different resolutions).*

8.4.2 $B_{u,d}^{**}$ Enrichment

$B_{u,d}^{**}$ are enriched in the following way: first, a neural network exploits the kinematical information attached to $B_{u,d}^{**}$ decays. The simultaneous treatment is possible because both mesons are assumed to have the same mass and both decay to B and π mesons. In the second step, the B_u^{**} and B_d^{**} mesons are separated after $B_{u,d}^{**}$ enrichment. This step is performed by the combination of the ChCorr network and of the HadId network. The idea behind the method is illustrated in figure 8.7. Having ideal separation power, which is of course not true in reality, the kinematically enriched sample can be subdivided into 4 different subsamples containing

- weakly decaying B^+ mesons, with **negative correlation** between the charge of the π and the flavour of the B^+. This sample, called the **right-sign B_d^{**} sample**, contains the $\mathbf{B}_d^{**}$ signal, decaying into $B_d^{**} \to B^+\pi^-$.

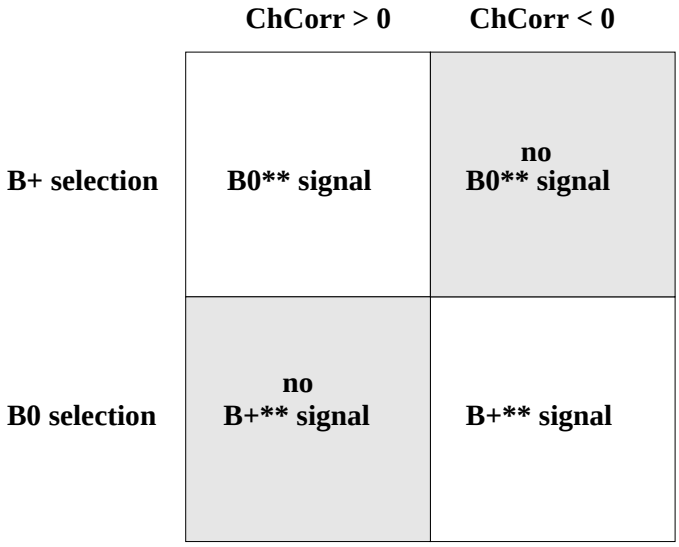

Figure 8.7: *The separation of B_u^{**} and B_d^{**} mesons after $B_{u,d}^{**}$ enrichment by purely kinematical information.*

- weakly decaying B^+ mesons, with **positive correlation** between the charge of the π and the flavour of the B^+. This sample, called the **wrong-sign B_d^{**} sample**, contains **no $\mathbf{B}_d^{**}$** signal.
- weakly decaying B^0 mesons, with **negative correlation** between the charge of the π and the flavour of the B^0. This sample, called the **wrong-sign B_u^{**} sample**, contains **no $\mathbf{B}_u^{**}$** signal.
- weakly decaying B^0 mesons, with **positive correlation** between the charge of the π and the flavour of the B^0. This sample, called the **right-sign B_u^{**} sample**, contains the $\mathbf{B}_u^{**}$ signal, decaying into $B_u^{**} \rightarrow B^0\pi^+$.

In reality, the subsamples contain certain levels of pollution by the other classes depending on the used cut values on the ChCorr and HadId networks. In the following the kinematical network is outlined and in the next section the results with the enriched samples are shown.

The Network Definition

The network with a single output node has been trained on the target $+1$ if the track was a pion originating directly from the B^{**} decay, and -1 for all other tracks. Tracks belonging to a hemisphere containing a B^{**} but not directly originating from the B^{**} decay are also regarded as background. The input variables presented to the network are closely related to the properties of the signal tracks. These are:

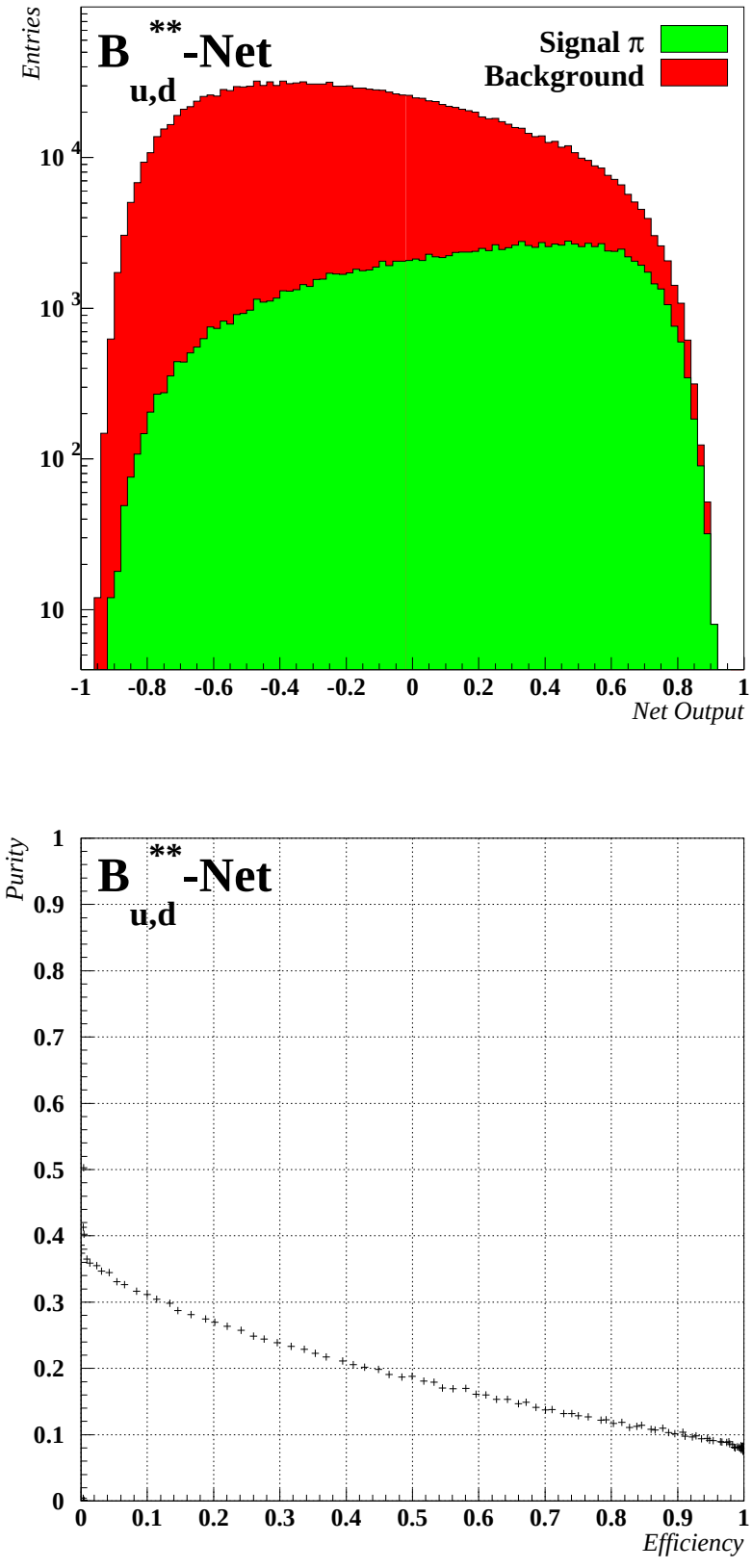

Figure 8.8: *The performance of the $B^{**}_{u,d}$-Net.*

- The **track rapidity, momentum** and the **energy weighted rapidity functions** for the B decay product as well as for the fragmentation and for the leading fragmentation hypothesis. In addition, the **rapidity** and the **momentum rank**.
- The **particle identification** network outputs for **electrons**, **kaons** and **protons**.
- The **pion decay angle in the B^{**} rest frame**. Signal pions are equally distributed, whereas background tracks are concentrated at backward angles with respect to the B^{**} flight direction.
- The laboratory opening angle α between the track candidate and the B flight direction.
- The **TrackNet** output and the **gap in the TrackNet** value between the two particles on either side of the TrackNet $= 0.5$ boundary.
- To estimate the quality of the vertex separation, the reconstructed **3D decay length** of the weakly decaying B hadron is given. The decay length corresponds to the best secondary vertex available in the hemisphere. In addition, the error on the decay length is also provided to the network together with the information on which vertex algorithm has been used (see 7.2.4).
- The simultaneous enrichment of B^+ and B^0 mesons is achieved by providing the **combined hemisphere** and **event b-tags** to the network and the B^+ and B^0 probabilities coming from the BHBN Network (Section 7.2.7). In addition, the two other BHBN output nodes, B_s and Λ_b are provided to the net as an anti-tag.
- Variables without separation power but with the ability to provide the network with quality information are the **hemisphere** and the **track quality** words of BSAURUS. As an indirect quality measure, the number of fragmentation tracks has been used.

Network Precuts

In addition to the precuts already included in the QCorr network, the following cuts, which are special for $B_{u,d}^{**}$ mesons, have been applied to enrich signal to background ratio. These cuts discard obvious background particles and ensure a better minimization performance of the network:

- The original Q-value estimate Q_0 is used to discard tracks with very high Q-value. The predicted Q-value for the narrow states is in the range of 200 to 300 MeV. The broad states should have approximately the same mass and abut width of about 200 MeV. Nevertheless, the

signal is expected to lie well under 1.0 GeV. This is taken to be the maximum value in the considered distributions.

- It has been found that all signal pions lie within the momentum range of 1.0 to 14.0 GeV, and tracks lying outside of this region can be considered as background. This has been checked on the special Monte Carlo sample to avoid cutting away regions possibly containing signal.

The performance of the $B^{**}_{u,d}$ Enrichment Network based on the kinematical variables mentioned above is shown in figure 8.8. The evaluation of the performance has been made on an independent Monte Carlo $q\bar{q}$ sample, with a natural track composition and after the application of the precuts.

8.4.3 Results for the $B^{**}_{u,d}$

The distribution of the measured $B^{**}_{u,d}$ Q-values obtained with the DELPHI detector is shown in figure 8.9. The plots are organized in the same way as in figure 8.7, i.e. the B^{**}_d are enriched in the upper left corner and the B^{**}_u are enriched in the lower right corner. The signal contributions to the upper right and lower left plots are due to the finite separation power of the networks. The networks that have been cut on to obtain the shown distributions in the data are: the charge correlation net Chcorr, the b-hadron type identification network HadId and the $B^{**}_{u,d}$ enrichment network. The specific cut on the ChCorr network has been

$$\begin{array}{rlll} \text{ChCorr} > 0.0 & B^{**0}_d \to B^+\pi^- & \text{right-sign } B^{**}_d\text{(upper left)} \\ \text{ChCorr} < 0.0 & B^{**0}_d \to B^+\pi^+ & \text{wrong-sign } B^{**}_d\text{(upper right)} \\ \text{ChCorr} > 0.3 & B^{**+}_u \to B^0\pi^- & \text{wrong-sign } B^{**}_u\text{(lower left)} \\ \text{ChCorr} < -0.3 & B^{**+}_u \to B^0\pi^+ & \text{right-sign } B^{**}_u\text{(lower right)} \end{array} \qquad (8.2)$$

.

The separation of the b-hadron types was performed by the cut on the B^+ and B^0 output nodes of the HadId networks. B^{**}_d enrichment was achieved by the combined cut on $P(B^+) > 40\%$ and $P(B^0) < 40\%$. Analogously, B^{**}_u enrichment was achieved by the combined cut on $P(B^+) < 40\%$ and $P(B^0) > 40\%$. The cuts have been chosen to match the output of the b-hadron enrichment networks shown in figure 7.13. Values above or below 40% correspond to hemispheres with enough information to tag or to anti-tag B^+ and B^0 mesons. The $B^{**}_{u,d}$ enrichment network has been cut to give an overall B^{**}_d purity of 30.7% in the B^{**}_d right-sign sample and an overall B^{**}_u purity of 27.1% in the B^{**}_u right-sign sample.

The contribution of B^{**}_d signal to the non-B^{**}_d-right-sign-plots is extracted to be 9% in the B^{**}_d wrong-sign sample, 5% in the B^{**}_u right-sign sample and 11% in the B^{**}_u wrong-sign sample. The percentage values are

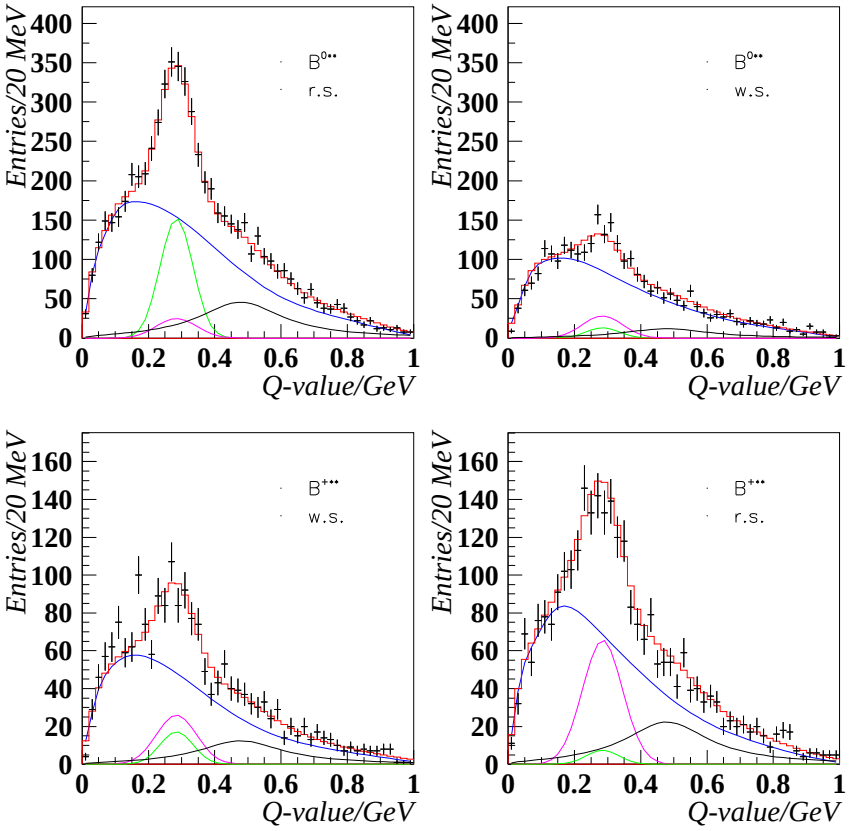

Figure 8.9: *Distribution of the Q-value of $B^{(*)}\pi$ pairs.*

evaluated by the number of B_d^{**} signal events in the different distributions in comparison to the number of signal events in the B_d^{**} right-sign sample. The contribution of B_u^{**} signal to the non-B_u^{**}-right-sign-plots is extracted to be 43% in the B_d^{**} wrong-sign sample, 38% in the B_d^{**} right-sign sample and 39% in the B_u^{**} wrong-sign sample. The ratio of B_u^{**} contamination in the non-signal samples is larger than the ratio of B_d^{**} contamination in its non-signal samples because of the worse B^0 enrichment performance. The B^0 are electrically neutral making its reconstruction more difficult and in addition it can mix with $\bar{B}^0$ mesons.

A combined fit to the four samples defined above was performed assuming that the observed signal is due to the narrow B^{**} states and that it could be described by a single Gaussian. A Breit-Wigner function was used to account for the additional broad feature observed in the data at about

500 MeV Q-value. In the following, the fit component to the narrow signal will be called *Gaussian* and the component to fit the broad signal will be called *Breit-Wigner* regardless of the actual physical states they are describing. The background shape is taken from Monte Carlo and smoothed. In the fit, the background is corrected as a linear function of the Q-value

$$\mathrm{BKG}_{MC} * (n + s \times (Q - 0.5)) \tag{8.3}$$

in units of GeV and where n and s are free parameters and can be different for the B_u^{**} and B_d^{**} spectrum. The constant 0.5 GeV subtracted off the Q-value is to reduce correlations between the parameters n and s. Without the subtraction, the two parameters are very strongly ($\approx 100\%$) negatively correlated, in addition to the their 'normal' correlation which is purely due to the data. Shifting the normalization point to the mean of a distribution removes to first order this artificial correlation.

The ratios of B_d^{**} and B_u^{**} in the r.s. sample compared to the other three samples are fixed to the values seen in the Monte Carlo. The free fit parameters are therefore:

- The **area** of the narrow B_d^{**} signal. The area of the narrow B_u^{**} signal is constrained to correspond to the same rate as of the B_d^{**} by taking into account their specific acceptance[3]. To fit the B_d^{**} area rather than the B_u^{**} area is an arbitrary choice in principle but the B_d^{**} provides a better signal because the B^+ net is superior to the B^0 enrichment net.

- The **Gaussian mean** $Q(B_{u,d}^{**} \to B^{(*)}\pi)$, fixed to the same value for B_d^{**} and B_u^{**}.

- The **Gaussian width** $\sigma_Q(B_u^{**})$ for the B_u^{**}.

- The **Gaussian width** $\sigma_Q(B_d^{**})$ for the B_d^{**}. The difference in the widths of the B_d^{**} and B_u^{**} spectrum is also observed in simulation and can be explained as a consequence of the worse reconstruction quality of B^0 mesons and of the worse B_u^{**} enrichment performance. It is not the effect of a different intrinsic width of the B_u^{**}.

- The **ratio R** between the broad and the narrow signal components. This ratio is constrained to be the same for both B_d^{**} and B_u^{**}.

- The common most probable value **MPV** of the additional Breit-Wigner functions (one for the B_d^{**} and one for the B_u^{**}). The same principle as above is applied to fix the rates of the B_d^{**} and B_u^{**} Breit-Wigner components to be equal.

[3] The rates are expected to be the same in any reasonable theory.

- The common full width at half maximum **FWHM** value of the additional Breit-Wigner functions.

The fit gives the following results:

$$\begin{aligned} Q(B_{u,d}^{**} \to B^{(*)}\pi) &= 285 \pm 3(\text{stat}) \pm 7(\text{syst})\text{MeV/c}^2 \\ \sigma_Q(B_d^{**}) &= 49 \pm 5(\text{stat}) \pm 6(\text{syst})\text{MeV/c}^2 \\ \sigma_Q(B_u^{**}) &= 60 \pm 5(\text{stat}) \pm 6(\text{syst})\text{MeV/c}^2 \\ R &= 0.8 \pm 0.2(\text{stat}) \pm 0.2(\text{syst}) \\ \text{FWHM} &= 574 \pm 90(\text{stat}) \pm 80(\text{syst})\text{MeV/c}^2 \\ \text{MPV} &= 480 \pm 20(\text{stat}) \pm 18(\text{syst})\text{MeV/c}^2 \ , \end{aligned} \tag{8.4}$$

obtained at the working point corresponding to figure 8.9. Figure 8.9 shows the result of a fit (histogram) together with the different signal and background contributions: the narrow B_d^{**} signal (green), the narrow B_u^{**} signal (magenta) and the sum of the Breit-Wigner components (black). Also shown is the contribution of the B_d^{**} to the B_u^{**} spectrum and vice versa. The values of the parameters in a scan series are shown in figures 8.10 to 8.14.

Production Rate

The production rate for the $B_{u,d}^{**}$ states can be extracted by correcting the Gaussian area for the acceptance which is calculated from the Monte Carlo. The results are normalized to the total number of selected hemispheres. The fitted rate of the Gaussian component is:

$$\frac{\sigma(B_{u,d}^{**})_{narrow}\text{Br}(B^{**} \to B^{(*)}\pi)}{\sigma_b} = 0.193 \pm 0.02(\text{stat}) \pm 0.01(\text{syst}), \tag{8.5}$$

i.e. approximately 20% of each b-jet contains a narrow B_u^{**} or B_d^{**} meson. In each step of the fit, the fitted value of the Gaussian rate is multiplied by the fitted value of the Gaussian to Breit-Wigner ratio in the same fit to obtain the rate measurement in that specific fit step. The broad rate is then based on the average over a whole scan procedure. The rate of the Breit-Wigner component at 500 MeV is:

$$\frac{\sigma(B_{u,d}^{**})_{broad}\text{Br}(B^{**} \to B^{(*)}\pi)}{\sigma_b} = 0.16 \pm 0.04(\text{stat}) \pm 0.04(\text{syst}) \ . \tag{8.6}$$

For the rate calculation it is assumed that 2/3 of all B^{**} decay into charged pions and 1/3 into (unobserved) π^0, according to isospin rules for an $I = 1/2$ decay into these mesons.

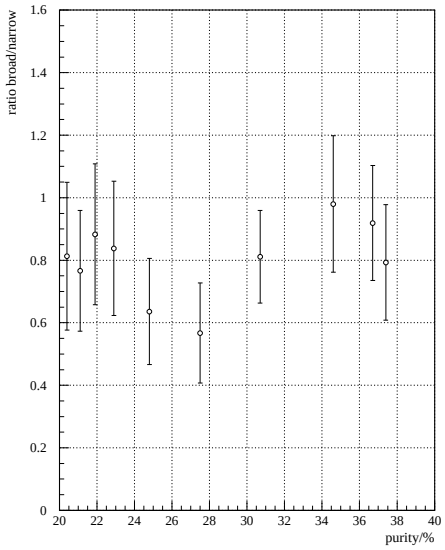

Figure 8.10: *Ratio of the Gaussian and Breit-Wigner components as a function of the $B^{**}_{u,d}$ purity.*

Statistical and Systematic Uncertainties

To estimate the systematic error attached to the measurement, a series of fits has been performed as a function of scanning over the B** purity by going to harder and harder cuts on the $B^{**}_{u,d}$ enrichment net. No check has been performed so far to the stability of the results regarding different models of background correction, e.g. comparing the linear correction applied so far to a correction function with more freedom, such as a quadratic correction.

To obtain the statistical error on the measurement, the mean of the statistical errors on the single points is taken. The systematic error is calculated to be the half of the difference between the largest and smallest values in a scan series. The scans are performed by cuts on the $B^{**}_{u,d}$ enrichment net to obtain measurements of the fit parameters at different purities and are shown in figures 8.10 to 8.14. Figure 8.10 shows the scan of the fitted ratio of the Gaussian and Breit-Wigner components. Figure 8.11 shows the scan of the fitted mean of the Gaussian (left-hand side) and of the most probable value of the Breit-Wigner (right-hand side). The fitted mean value of the Gaussian component has a significant trend as a function of $B^{**}_{u,d}$ purity. This behaviour is perhaps due to an incorrect modelling of the background or due to an incorrect ansatz for the signal and is a source of a residual systematic. A less naive signal shape could possibly account for the

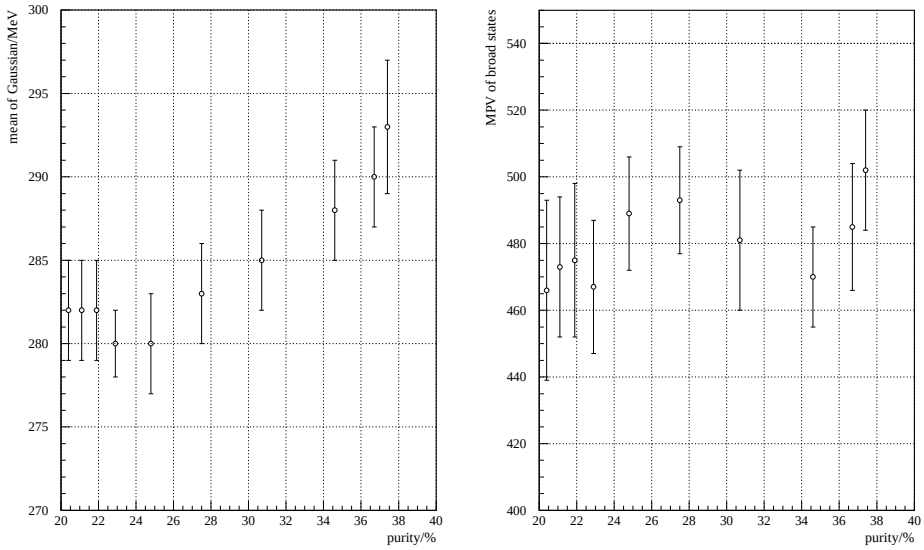

Figure 8.11: *A* **mass** *scan over $B_{u,d}^{**}$ purities at different enrichment net cuts for the Gaussian (left) and Breit-Wigner (right) fit components.*

bias. Figure 8.12 shows the scan of the fitted Gaussian rate (left-hand side). On the right-hand side, the evaluated rate of the Breit-Wigner component is shown. The rate of the Breit-Wigner component is calculated from the Gaussian rate multiplied by the fitted Breit-Wigner to Gaussian ratio. The plots in figure 8.13 show the fitted widths and FWHM of the Gaussian and of the Breit-Wigner. The plots on the left-hand side and in the middle correspond to the width of the B_u^{**} (left) and B_d^{**} (middle) mesons. The Gaussian widths show, like their means, a trend as a function of $B_{u,d}^{**}$ purity. Again, this can be caused by an improper modelling of the background or by a too naive treatment of the signal components. The spread in the fitted values is a dominant contribution to the systematic error on the widths of the B_d^{**} and B_u^{**} Gaussian components. Figure 8.14 shows the normalization and slope correction parameters for both B_d^{**} and B_u^{**} enriched samples. The parameters have a significant tendency reflecting the complex nature of the so called background. The background, at least as it is defined in this fit, may be a composition of real B^{**} background and additional broad states not accounted for in the fit. The tendency in the fit will remain until these components are well understood and under control. Also the modelling of the background coming from the fragmentation process may be not ideal building another source of uncertainty in the spectrum.

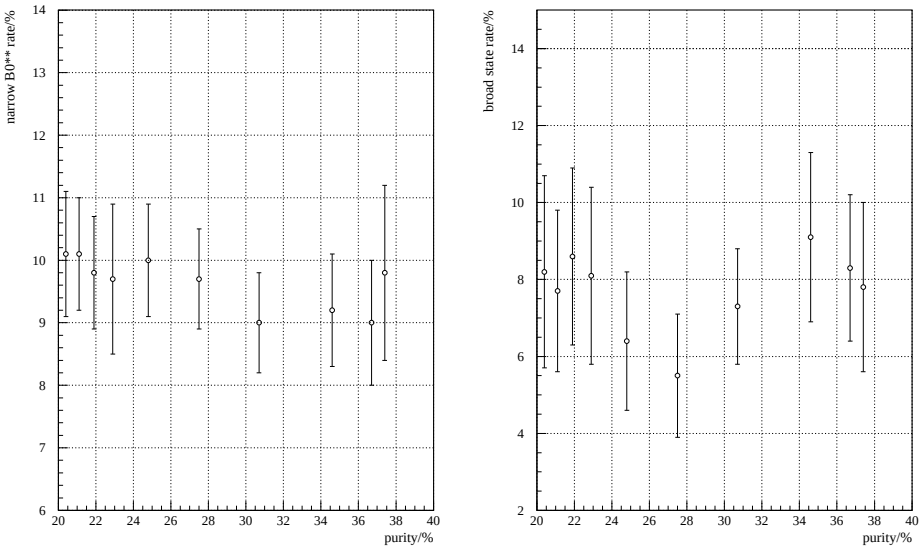

Figure 8.12: *A* **rate** *scan over $B^{**}_{u,d}$ purities at different enrichment net cuts for the Gaussian (left) and Breit-Wigner (right) fit components.*

8.4.4 Interpretation and Outlook

The interpretation of the results in the framework of HQET is very dependent on the assumption going into the fit. The fit performed above has been based on the assumption that the data comprises two main features: a narrow peak at 290 MeV and a broad peak at approximately 500 MeV. In the fit procedure though, no constraints were made to force the fit to the observed Q-values of the resonances. Making the classical approach of spectroscopy, the narrow peak is fitted by a Gaussian function assuming a very narrow intrinsic width of the underlying state. The observed width is thus not the width of the state but is due to the finite detector resolution. The broad resonance can be naturally described by a Breit-Wigner function, which is not distorted by the resolution due to its large width (FWHM). In the fit to the data, the Breit-Wigner tends to the broad peak and the Gaussian to the narrow peak.

The interpretation of the broad peak is not clear. A possible interpretation is that it corresponds to a broad B^{**} state, which would mean that spin-orbit inversion is valid (see section 3.7), i.e. the broad states appear at higher masses than the narrow states. It can also not be ruled out that the Breit-Wigner is a radial excitation. HQET predictions state a Q-value between 400 and 500 MeV for the radial states B', e.g. [EGF98] and their

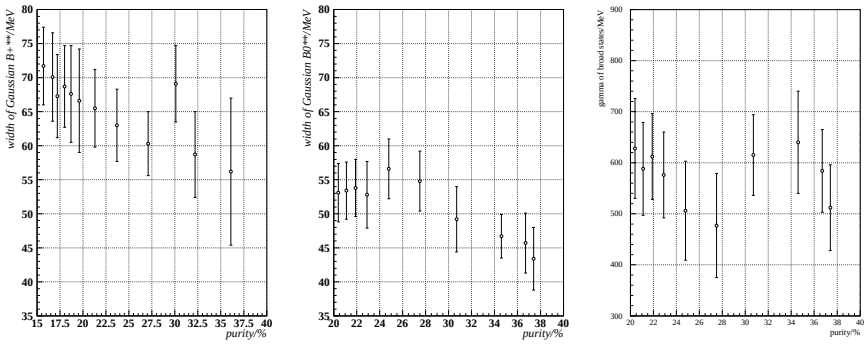

Figure 8.13: *The plots show the scan over the attained purities at different $B_{u,d}^{**}$ enrichment net cuts for the* **widths** *and* **FWHM** *of the Gaussian and Breit-Wigner (right) fit components. The widths of the B_u^{**} (left) and B_d^{**} (middle) have been fitted individually thus accounting for the different reconstruction quality of B^+ and B^0 mesons.*

widths are commonly believed to be in the range 100 to 200 MeV.

The narrow component has a width which varies in the range 43 to 53 MeV for the B_d^{**} and in the range 59 to 69 MeV for the B_u^{**}. A comparison of this result to the anticipated resolution of 33 MeV for the B_d^{**} and of 36 MeV for the B_u^{**} at this Q-value shows that the assumption of a single narrow resonance as a source of this peak has to be given up. A possible contribution to this peak would be the decay of the 1^+ meson to $B^*\pi$ and of the 2^+ meson to $B^*\pi$ and $B\pi$. Taking into account the 46 MeV mass difference between the $B^*\pi$ and $B\pi$ decay channels of the B_2^* and the predicted mass splitting of 12 MeV between the B_2^* and B_1 states, the overall width of the total signal would be of the order of 35 MeV assuming a $1^+ : 2^+$ production ratio of 3:5. As a conclusion, the narrow signal component of the data does not rule out the HQET mass splitting prediction. If the observed width is due to the narrow B^{**} states only, the mass splitting could be as high as 50 MeV. To evaluate this, Gaussian functions with relative areas fixed to the HQET prediction of the production and decay rates of the B_2^* and B_1 states are added in a histogram. Fixing the mass splitting between the $B^*\pi$ and $B\pi$ decays of the B_2^* to 46 MeV, the hyperfine splitting between the 1^+ and the 2^+ states was increased until the total width of the three added Gaussians became 50 MeV (equal to the fitted width of the Gaussian component in the data) obtained by a single Gaussian fit.

As already stated, it cannot be ruled out that the feature at $Q \approx$ 500 MeV is due to a radial excitation state. In fact, HQET does not favour orbitally excited broad states at this Q-value. Assuming the radial state, the broad B^{**} states could perhaps lie underneath or at least in the proxim-

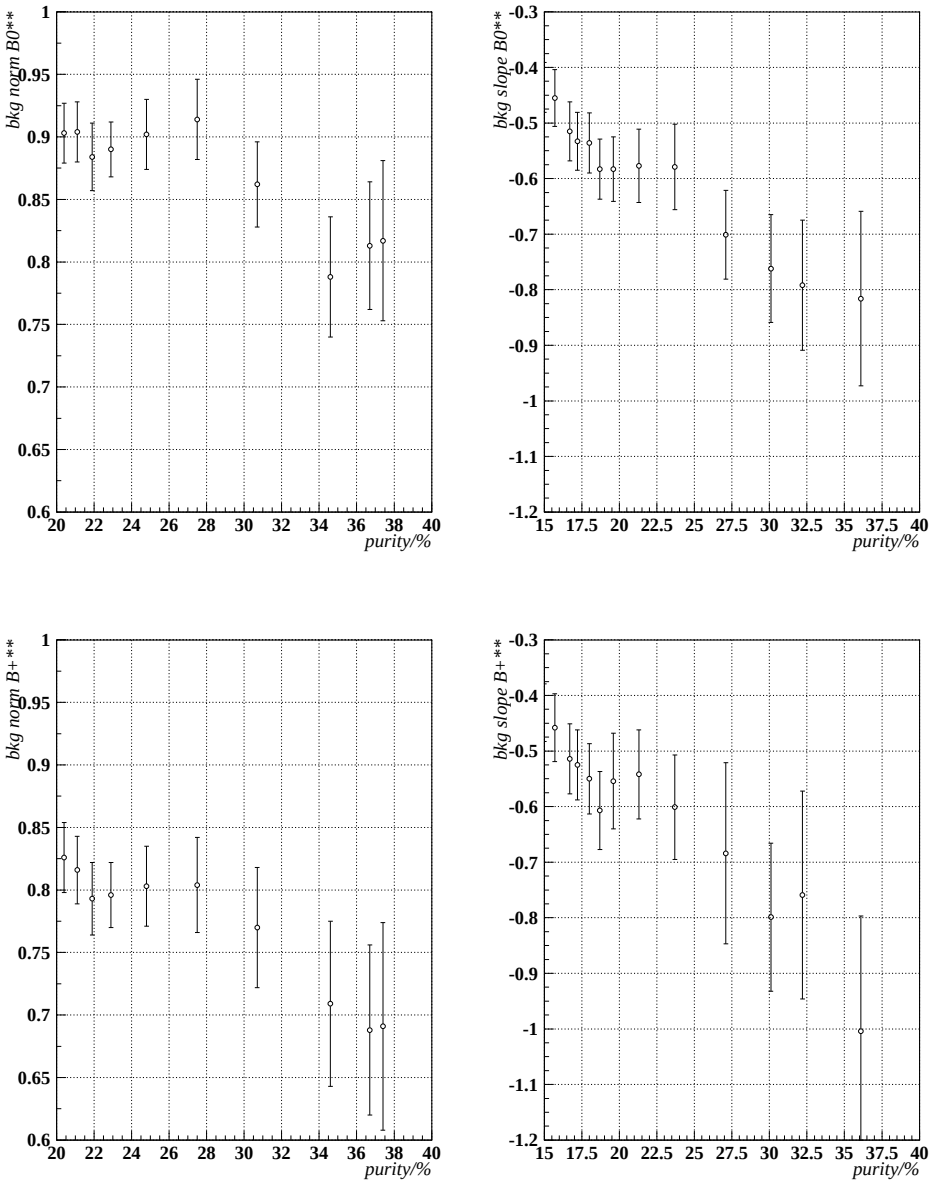

Figure 8.14: *Scan of the $B^{**}_{u,d}$ normalization and slope correction parameters (top) for the B^{**}_{d} enriched samples and (bottom) for the B^{**}_{u} enriched samples.*

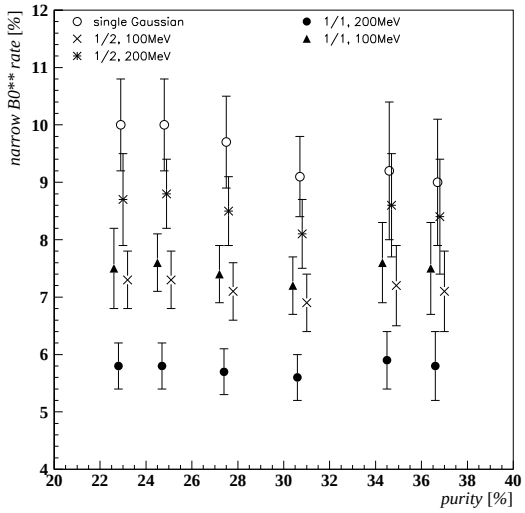

Figure 8.15: *The effect of adding another signal component to the fit in the proximity of the first Gaussian.*

ity of the narrow states. Fitting an additional Gaussian component to the data just underneath the narrow peak will naturally reduce the measured rate and width. In order to demonstrate this, another series of fits has been made with an additional Gaussian centred at the same Q-value as the first (narrow) Gaussian but with a fixed area and width. A scan was performed over various choices of this width. The idea was that this second Gaussian would roughly account for any broad component to the signal occurring at Q-values around the narrow peak position. The result for the 'narrow' rate is presented in figure 8.15 where the width of the 2nd Gaussian was varied between 100 and 200 MeV and the ratio of the first to the second Gaussian, i.e. 'broad' to 'narrow' was varied between the expected limits of 1:1 (state counting) to 1:2 (spin counting). The rate of the narrow B^{**} states can clearly be made to come down significantly if such a broad component is accounted for and without it the 'narrow' rate measured is most probably some mixture of (narrow+broad) rate. The total rates obtained in the process were compatible to the rate measurement with just one Gaussian and the χ^2 of these fits was within 3σ, i.e. the results were indistinguishable.

As already mentioned earlier, the tendency seen in figure 8.14 concerning the correction parameters of the background component reflects the uncertainty in the decomposition of the signal and also in the optimal treatment of the fragmentation process dominantly accounting for the background shape. The results on the rate of the narrow and broad components at 280 and

500 MeV respectively are stable thus allowing to confirm the existence of the narrow $B^{**}_{u,d}$ states, though without being able to separate them due to resolution restrictions, and to observe that there is evidence for a broad signal component possibly due to a radial excitation.

As an outlook to the future it can be stated that these components have to be better understood before the analysis can be called completed. In particular, the fit ansatz with only one Gaussian should be altered to a more sophisticated one comprising the expected narrow states convoluted with the detector resolution. Very strong cuts on the enrichment networks could be the basis for a helicity study of the narrow states thus confirming or disproving the assumption of their J^P assignments. As explained, there has been a quick check concerning the broad states underneath the narrow ones. A detailed study concerning these states with an appropriate choice for their parametrization (Breit-Wigner) would allow for less tentative results. Also a better understanding of the background and a more flexible correction function are essential for the successful completion.

8.5 A Study of the B^{**}_s

This section reports on a study of orbitally excited strange B mesons in the decay mode $B^{**}_s \to B^{(*)+}K^-$. If the B^{**}_s mass is above the $B^{(*)}K$ threshold, this will be the dominant decay mode since the channel $B^{(*)}_s\pi$ is forbidden by isospin conservation. The decay into $B^{(*)}K^0$ is difficult to observe experimentally due to the limited K^0 reconstruction efficiency and to the dominant K^0 background from B decays. Therefore, the search focuses on **narrow** B^{**}_s states decaying into charged particles. Again, as for the $B^{**}_{u,d}$ mesons no distinction is made between decays into BK and B^*K and thus a shift of 46 MeV in Q-value is expected for the latter one.

8.5.1 The Q-Value

The Q-value of the decay is evaluated by a dedicated QCorr neural network, which is optimized for B^{**}_s decays. The training of the B^{**}_s QCorr network has been performed on a special Monte Carlo sample, which has a flat Q-value distribution for B^{**}_s decays. In addition to the precuts mentioned before in section 8.3.1, the momentum of the accepted tracks was limited to the range 1.0 to 10.0 GeV. This is the anticipated momentum range of the signal kaons. The training has been performed on a sample of approximately 120.000 signal patterns after precuts. The Q-value correction ΔQ was evaluated via the maximum likelihood estimator of the probability density distribution and the error on the Q-value as half of the 68% confidence interval around its median. The obtained track-by-track resolution estimate σ_Q is shown in figure 8.16 as a function of the reconstructed Q-value.

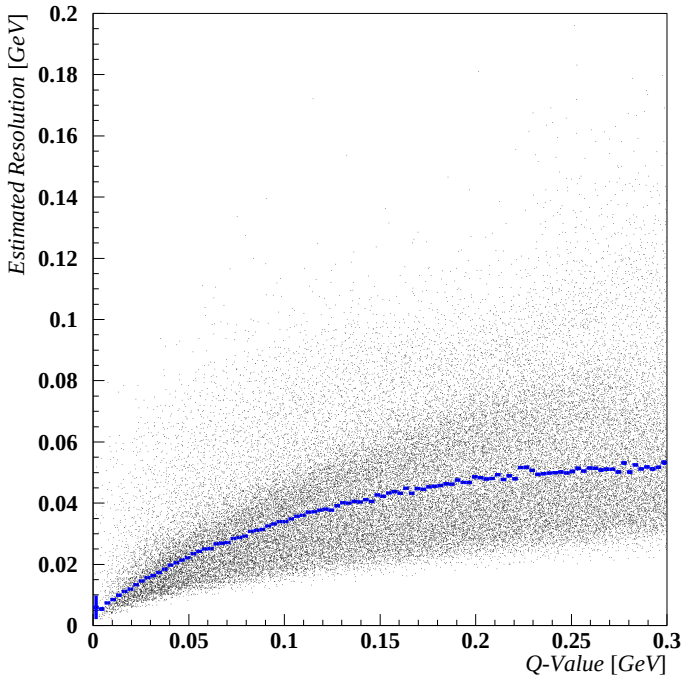

Figure 8.16: *Median of the Q-value resolution σ_Q provided by the B_s^{0**} QCorr Network as a function of the measured Q-value.*

8.5.2 B_s^{**} Enrichment

The second key element of the analysis of B_s^{**} states is the enrichment of the Q-value distribution in signal events. The method used is the neural network ansatz described in section 8.3.2. In contrast to the enrichment of B_u^{**} and B_d^{**} mesons, the available information is not applied in steps to achieve higher and higher purities but all information is fed into one single network. This is, first of all, feasible in the case of B_s^{**} mesons (it was not for $B_{u,d}^{**}$ mesons) and, second, the network, which is trained in this way, has a better performance since it can also consider the possible correlations. To repeat again, the informations available for signal enrichment are: kinematic correlations between B and K, charge correlations between B and K, the b-hadron Type Identification Network and, very important, the Kaon Identification Network MACRIB.

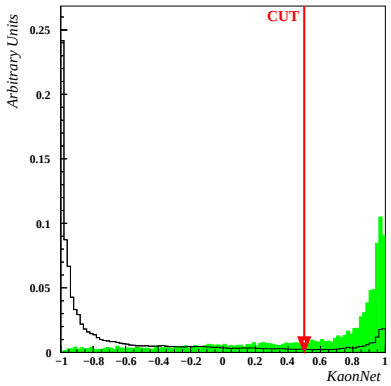

Figure 8.17: *The KaonNet distributions of B_s^{**} signal and background tracks. A relative hard cut on the KaonNet output at KaonNet > 0.5 is applied prior to network training to achieve a kaon purity of $\sim$67% and an efficiency of $\sim$73%.*

Network Definition

The B_s^{**} neural network is trained to classify kaon tracks originating from the B_s^{**} decay. The network has two main tasks to perform. Firstly, a separation of kaons from pions, which form the main source of background, and secondly a separation of signal kaons from non-signal kaons. Background kaons at high purities originate largely from leading fragmentation particles, produced right after the primary hadron, with a very high rapidity, almost like signal kaons. Besides the kinematical variables that go into the $B_{u,d}^{**}$ network, the following variables reflecting the characteristics of hemispheres containing B_s^{**} mesons are:

- The KaonNet output of the first three leading fragmentation particles. The sorting is performed by their rapidity value, and the decision if a particle is originating from the fragmentation or B decay is made by the particle's TrackNet value. If this is below 0.5, it is considered to be a fragmentation track. The production of B_s^{**} mesons in the fragmentation process involves the production of an additional $\bar{s}$ quark in the vicinity of the primary meson. The first to third hadron in the fragmentation following the B_s^{**} has an s quark content and is or decays into a kaon. In the string fragmentation picture, the rank of the produced particles, i.e. if they e.g. represent the leading or the next-to-leading fragmentation particle, is strongly correlated to their rapidity. Since it is unknown which of the produced fragmentation particles contains the $\bar{s}$ quark, the first three particles in the rapidity rank are taken.

- Analogous inputs provide the network with the information on the estimated number of charged kaons originating from fragmentation and from B decay.

- Reconstructed K_s^0 are counted if their rapidity is above 1.6 and their number is added as an input to the network.

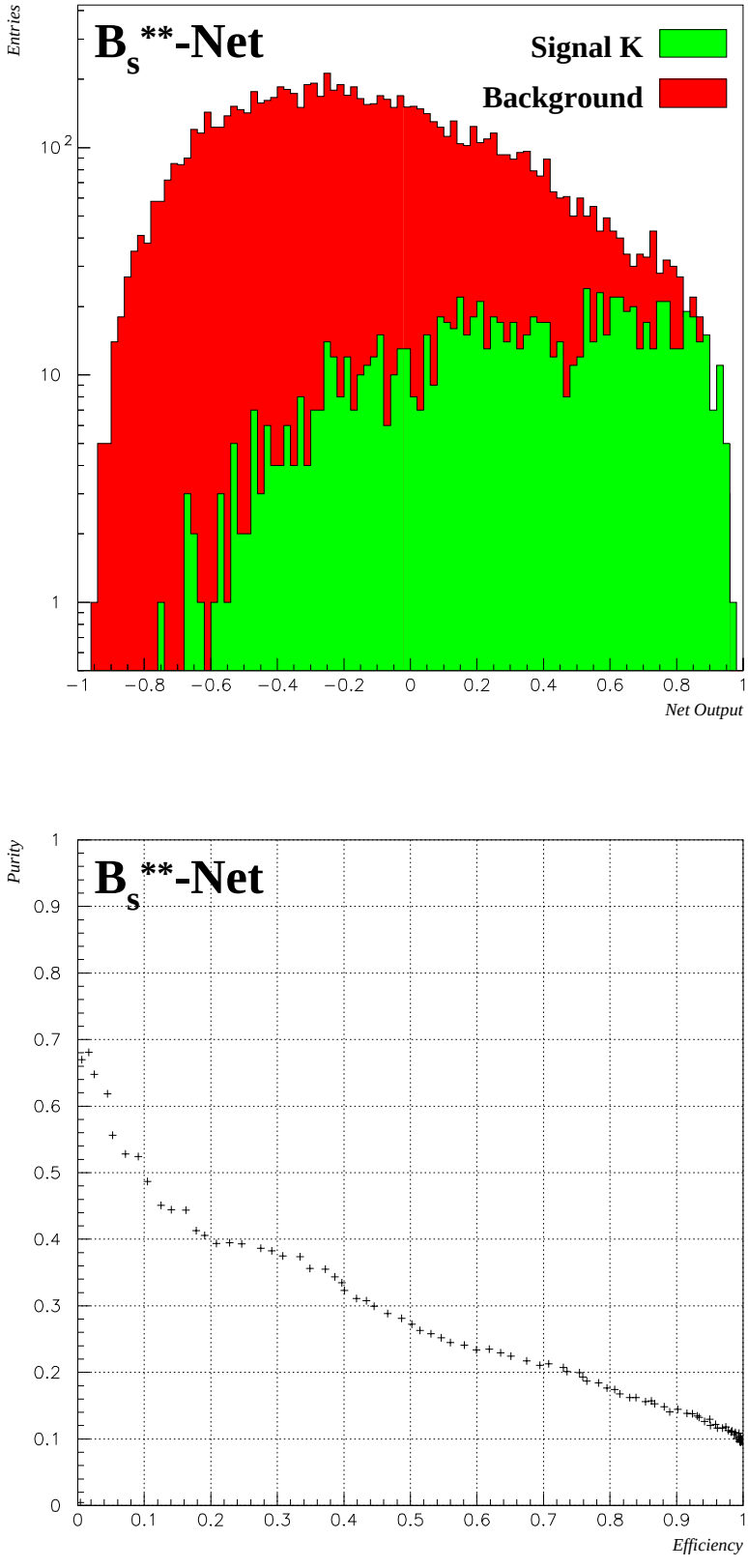

Figure 8.18: *The performance of the B_s^{0**}-Net.*

In addition to the purely kinematical input variables, the variables listed below provide the network with a correlation between the B flavour and the charge of the track. The further input variables, multiplied by the track charge, are:

- The single hemisphere B^+ decay flavour (DFLBP) net combined with the single-sided production flavour net of the other hemisphere. This has the advantage of excluding the B^{**} track candidate from the determination of the B^+ decay flavour, since the DFLBP net does not include fragmentation tracks.
- The jet charge minus the jet charge of the opposite hemisphere.
- The number of pions weighted by their charge and the probability to come from behind or in front of the secondary vertex. For example, for the $B^+ \to D^-\pi^+\pi^+$ decay one expects positive pions behind the secondary vertex and negative ones in front.
- The both hemisphere decay flavour net.
- The likelihood-ratio-based production flavour tag for B_s and B^0.

The performance of the B_s^{**} enrichment network is shown in figure 8.18. Again, the performance has been evaluated on an independent $q\bar{q}$ Monte Carlo sample with a natural composition and after the application of the precuts.

The Network Precuts

Since the rate of B_s^{**} mesons is rather small and only one out of all tracks in a hemisphere is the signal kaon, the track-level signal purity is approximately 0.16%. The training, with no cuts applied, leads to a network with a low enrichment power. This results from the fact that the kaon net is highly correlated with the desired B_s^{**} net output and at the same time pions dominate the training sample. The network, then, mainly separates pions from kaons which sometimes originate from the B_s^{**} decay. This means that the second task of the net, namely to separate signal from background kaons, has not been performed. To achieve this, a relatively hard cut on the MACRIB kaon net has to be applied (see figure 8.17).

Since signal kaons have a momentum between 1 and 10 GeV, tracks outside of this momentum window are not considered in the analysis. The Q-value of both the broad and the narrow states is expected to lie well below 400 MeV, and this value is taken as the maximum for the Q-value plots. This allows to cut on tracks with a Q-value measurement above this maximum. Beside these special cuts, the standard network precuts (see figure 8.5) have been used for the training. The resulting track-level signal purity is of the

level of 9%, as evaluated in the standard DELPHI $q\bar{q}$ Monte Carlo sample. This purity, though, is dependent on the rate put into the generator. It has been proven that the training on this reduced sample leads to a better signal purity, due to the separation of the kaon sources.

Reflecting the extremely sensitive performance of the B_s^{**} Enrichment Network to the precut applied on the KaonNet three networks have been trained with different kaon purities in the training sample. The kaon purity levels were approximately: 42%, 67% and 80%. In figure 8.18 only the performance of the network with the KaonNet precut giving 67% purity is shown. The networks with 42% and 80% kaon purity are used to provide a cross check in the rate measurement and to build the basis of systematic uncertainty studies.

8.5.3 Results for the B_s^{**}

A B_s^{**} enriched Q-value distribution is shown in figure 8.19. To obtain the distribution, a cut on the B_s^{**} Enrichment Network (with kaon purity 67%) has been applied. To achieve a good resolution of the signal, only the 75% best quality tracks have been chosen. The information on the quality is provided by the QCorr network as shown in figure 8.16. Each candidate entering the distribution is assigned the optimal Q-value along with its error σ_Q. By determining the 75% quantile of the scatter plot in figure 8.16, one can select candidates having a better resolution assigned to them as the 75% quantile. This way, bad quality candidates dominantly contributing to a degraded Q-value resolution do not enter the distribution. With the applied B_s^{**} network cut in combination with the quality cut, a B_s^{**} purity of $\approx 22\%$ is attained. This purity level is given by the standard DELPHI Monte Carlo for generic $q\bar{q}$ events.

The excess in data over the Monte Carlo is compatible to detector resolution at its Q-value (≈ 80 MeV). The expected signal, as predicted by HQET, is a combination of resonances originating from the four states of the B_s^{**} meson (see chapter 3). This fit ansatz performed on the data makes the assumption that the visible resonance is due to a single resonance only. The theoretical expectation of the widths of the narrow B_s^{**} states are of the order of an MeV making the fit ansatz with a single Gaussian a plausible choice since the width of the measured peak is then due to detector resolution only. In addition to the Gaussian fit component, the background is fitted by the background given in the Monte Carlo linearly corrected for its normalization and slope by applying

$$\mathrm{BKG}_{MC} * (n + s \times (Q - 0.15)) \ . \tag{8.7}$$

Altogether, the aim of the fit is to establish a stable result, i.e. a stable measurement of the mean and rate of the peak, and as a consequence to prove the existence of the B_s^{**} resonance as such. Details concerning the

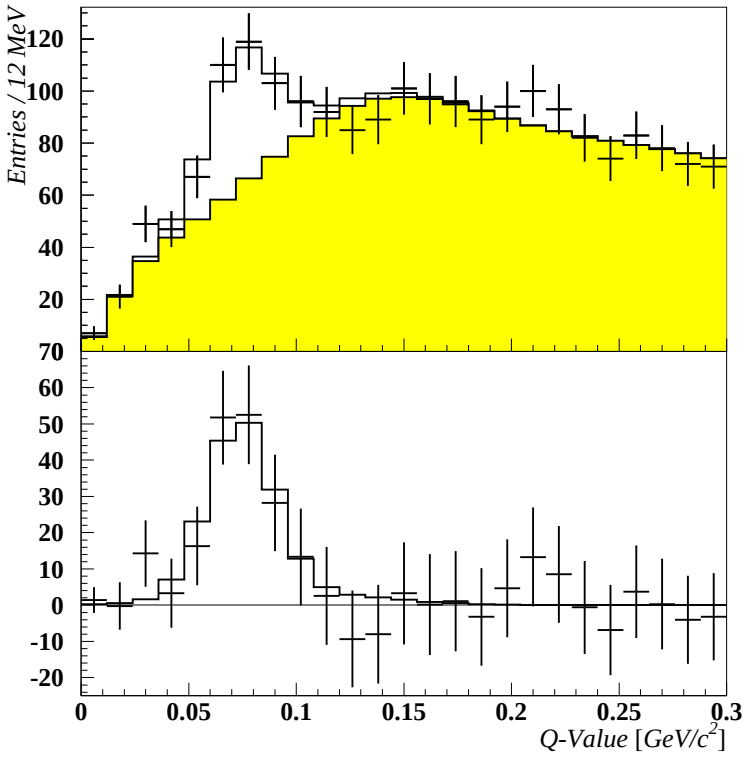

Figure 8.19: *Above, the distribution of the Q-value of $B^{(*)}K$ pairs (data points) is shown along with the Monte Carlo expectation of the background (shaded area). Below, the background subtracted Q-value distribution is shown. The resonance was fitted with the expected detector resolution to give the solid curve.*

splitting of the narrow states and their production and decay ratios are not tackled, neither is the aim to prove the existence and the properties of the broad states. The broad states are explicitely not considered in the chosen fit ansatz since they are automatically absorbed by the background and its linear correction function. The free parameters of the fit are:

- The **area** of the narrow Gaussian.
- The **Gaussian mean** $Q(B_s^{**} \to B^{(*)}K)$. The fitted value of the mean is not constrained to the observed (by eye) Q-value of the excess. The Gaussian mean could take on any value within the fitted histogram.
- The **Gaussian width** σ_Q. The width is also **not** constrained to any range. A possible constraint would be to limit the minimum value to

the median resolution at the position of the Gaussian mean.

- The background normalization correction **n** and the background slope correction **s**. Also these parameters are not constrained.

Figure 8.20 shows the fit result of the free parameters and the calculated rate measurement. The rate is the acceptance corrected signal area divided by the total number of b-hemispheres[4]. The entries to the various histograms are the result of nine consecutive fits performed on data samples with increasing B_s^{**} purity. The area of the Gaussian decreases as the purity increases due to efficiency losses (top left of figure 8.20). The mean and width of the Gaussian signal component are very stable as a function of the enrichment. The background normalization correction is stable but not stable in its slope correction thus showing that with increasing B_s^{**} purity, the assumption of having only a single narrow signal peak and **pure** background is not valid any more. Due to the chosen fit ansatz, this is not a surprising result. The fits concerning the narrow resonance are nevertheless good and provide compatible results with the fits at low purities.

The mean fitted width of the free Gaussian - free since its width is not constrained by any means - is approximately 11 MeV. Comparing this result to the median Q-value resolution $\sigma_Q(Q)$ at the measured Gaussian mean 75.6 ± 0.9 MeV gives $\sigma_Q(76\text{ MeV}) \approx 17$ MeV (provided by a histogram similar to the one in figure 8.16 but with the apropriate cuts applied) shows a slightly too narrow result. Though the stability of the rate measurement (see figure 8.20) indicates that the peak is **not** a statistical fluctuation, the narrowness of the signal was an initiator of a further fit procedure taking into account the full expected detector effects on a possible narrow signal.

To estimate the detector effect on a narrow resonance, the fit with the free Gaussian signal component was taken as a basis. As mentioned in section 8.5.1, for the training of the B_s^{**} QCorr network and for the training of the B_s^{**} enrichment network, a dedicated Monte Carlo sample was created with the mass of the B_s^{**} meson spread over many hundred MeV. The Q-value distribution of this special simulation is thus flat. A subsample of the produced B_s^{**} mesons with generated Q_{true}-values between 80 ± 10 MeV was selected. The effect of the detector was modelled by applying the same Q-value reconstruction algorithms on the selected sample as used for the reconstruction of the data. To obtain the expected signal shape at 80 MeV Q-value, the residual Q-value ($Q_{measured} - Q_{true}$) was plotted. An exemplary residual distribution is shown in figure 8.21 fitted with a double Gaussian, which has proven to be an appropriate parametrization at all purities. The two Gaussians divide the residual Q-value distribution into two samples: a

[4] In the shown fit result, the rate measurement is valid for B_s^{**} decaying into B^+K^-. The final experimental result of the rate includes the decays into the neutral final state B^0K^0 also, thus being twice as high as the values in the scan plot.

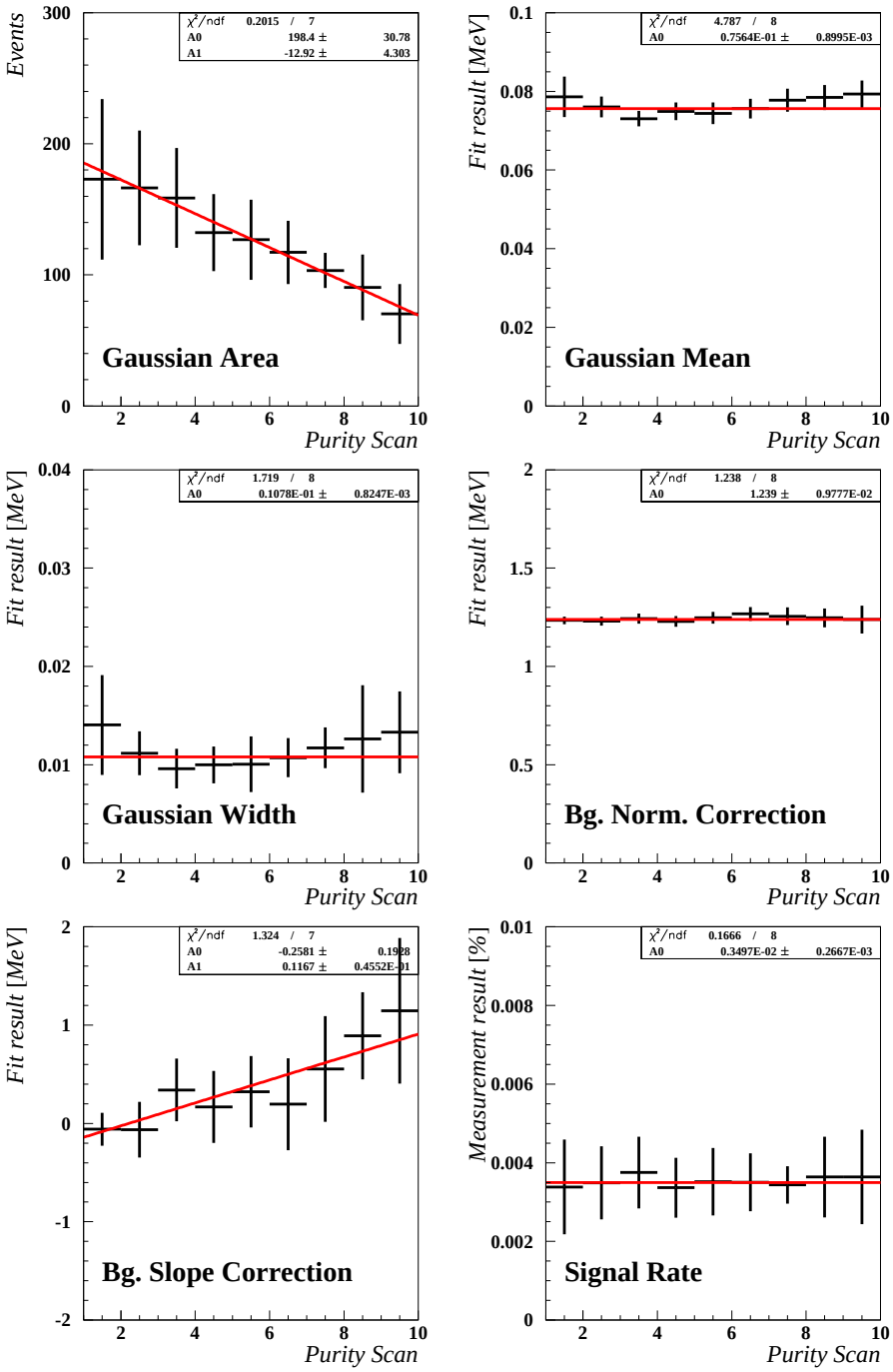

Figure 8.20: *The free fit parameters of the B_s^{**} Q-value distribution along with the calculated narrow B_s^{**} rate.*

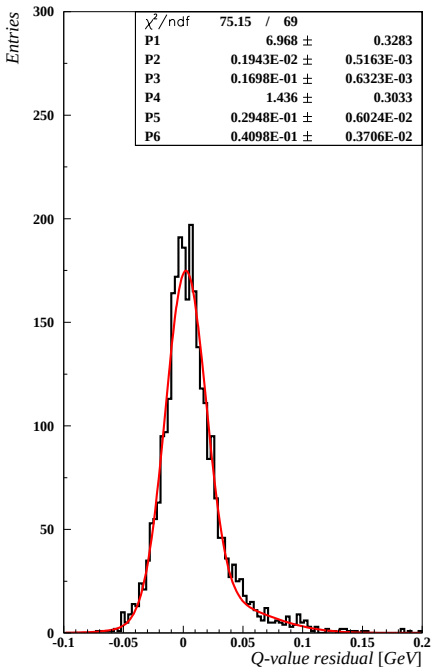

Figure 8.21: *The Monte Carlo residual Q-value distribution for the case a narrow B_s^{**} state was produced at 80 MeV Q-value. The fit parameters (area,mean and width) belong to two Gaussians.*

good quality sample which is narrow and a bad quality sample which is broad thus describing the tails of the Q-value resolution. The fitted result, more precisely the parametrized function, was taken as the signal component to be fitted to the data instead of the single free Gaussian of the first attempt. The expected signal shape was automatically evaluated in the described way for each considered B_s^{**} purity and for each cut on the quality, i.e. resolution, of the BK candidates. The fits, i.e. the free Gaussian fit and the fit with the expected signal shape, gave both very good χ^2 results (between 40 and 120 with $n_{dof} = 96$ and 95) with very low $\Delta\chi^2$ differences.

To obtain an optimal result, the signal component was allowed to be shifted as a whole in Q-value and to be adjusted in its area. In the procedure, the relative areas of the two Gaussians describing the signal shape were kept the same as well as their relative distance and widths. The definition of the background component has not been altered. The free parameters of the fit are:

- The **area** of the signal shape.
- The **mean** of the signal shape defined by the position of the narrow Gaussian component of the fit to the residual Q-value distribution. This fit parameter corresponds to the measured Q-value of the decay: $Q(B_s^{**} \to B^{(*)}K)$. The value of the mean parameter was not

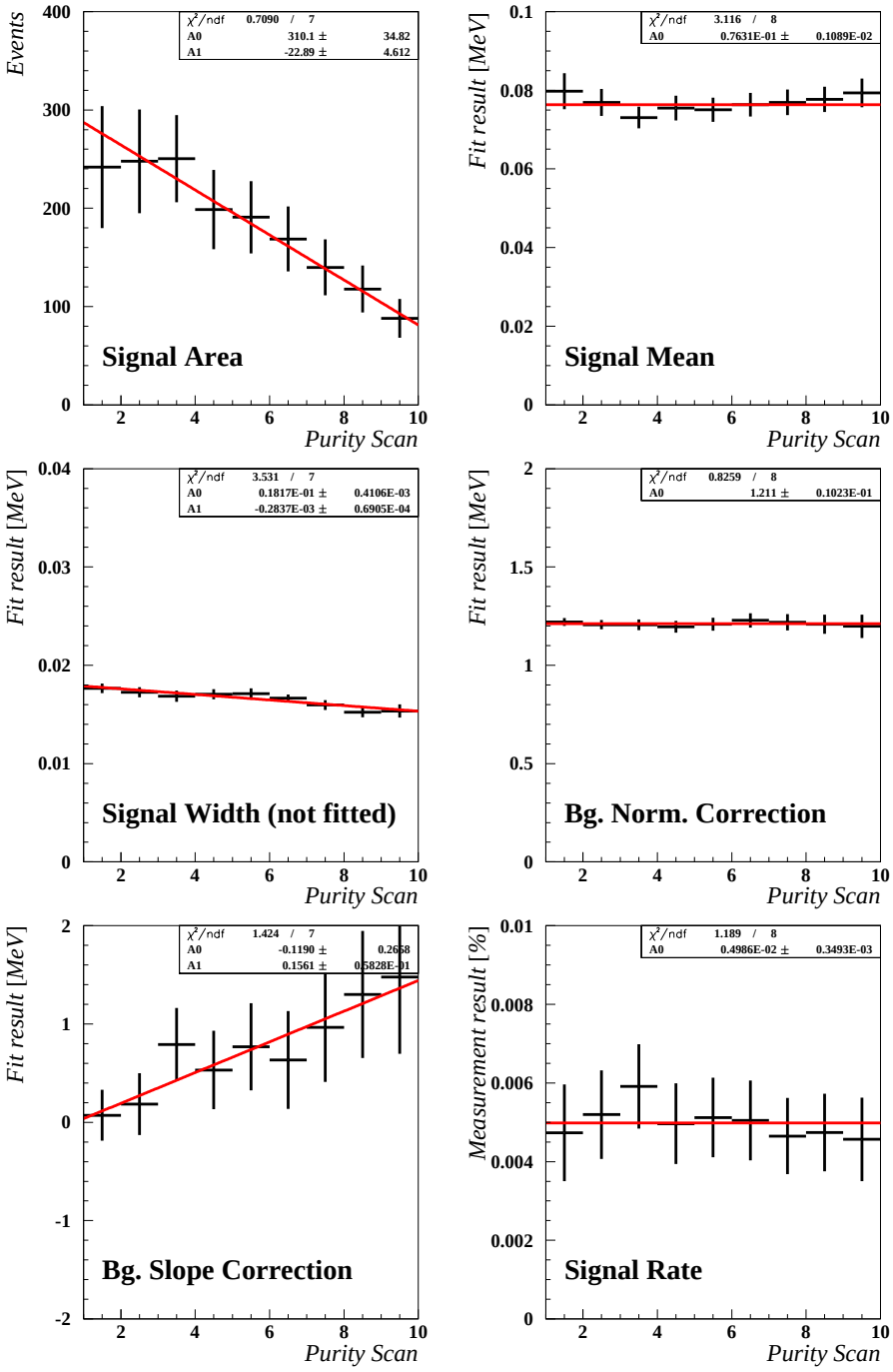

Figure 8.22: *The free fit parameters of the B_s^{**} Q-value distribution along with the calculated narrow B_s^{**} rate and the width of the narrow Gaussian component of the fit to the residual Q-value distribution.*

constrained so that it could take on any value within the fitted histogram, though a large deviation would have deteriorated the result since the shape of the signal is only valid at 80 ± 10 MeV Q-value due to the strong dependence of σ_Q from Q.

- The background normalization correction **n** and the background slope correction **s**. Also these parameters are not constrained.

The free fit parameters are shown in figure 8.22. Also shown in the figure is the width of the narrow Gaussian component of the signal shape parametrization. The significant slope on the signal 'width' is due to the better Q-value resolution at higher and higher B_s^{**} purities. The reason is that in hemispheres with a relatively large B decay length and with a relatively well separated K from the secondary vertex, an optimal Q-value resolution can be achieved and at the same time an optimal identification of B_s^{**} is possible. The mean narrow width of the signal shape is $\sigma(Q) = 16.7$ MeV. The average over all performed fits gives for the Q-value

$$Q(B_s^{**} \to BK) \quad = \quad 80.2 \pm 4.7(\text{stat}) \pm 4.2(\text{syst})\text{MeV/c}^2 \tag{8.8}$$

corresponding to the mass

$$Q(B_s^{**}) = 5853.2 \pm 4.7(\text{stat}) \pm 4.2(\text{syst})\text{MeV/c}^2 \ . \tag{8.9}$$

This result is obtained by a scan over all variables contributing to the systematic error quoted. It represents the mean of the Gaussian fitted to the collected sample of measurements. It is **not** the Q-value of the signal shown in figure 8.19 and it is also **not** the result of a chosen working point.

On the lower right side of figure 8.22, the corresponding rate measurement is shown, which is evaluated by the acceptance corrected signal area divided by the total number of selected b-hemispheres. Taking into account the Clebsch-Gordan coefficients for the decay of B_s^{**} mesons to the charged and neutral BK channels (50%:50%), the rate based on the area of the fitted signal peak is

$$\frac{\sigma(B_s^{**})\ \text{Br}(B_s^{**} \to BK)}{\sigma_b} \quad = \quad 0.0085 \pm 0.0023(\text{stat}) \pm 0.0013(\text{syst}) \ , \tag{8.10}$$

i.e. in almost 1% of all b-events a narrow orbitally excited B_s meson is produced. This rate measurement is $\approx 20\%$ higher than the rate measurement which would be obtained by the single Gaussian fit with free width. The difference can be accounted for by the additional contribution of the tail in the signal shape. Still, the results are compatible and so represent a proof that the measured peak using the single Gaussian fit ansatz was not a background fluctuation although its width was smaller than the detector resolution estimated by the QCorr network.

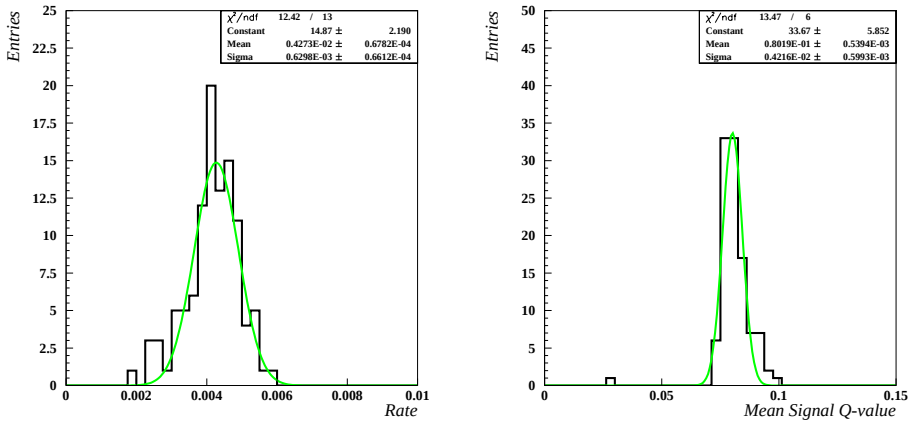

Figure 8.23: *Left: The $B_s^{**} \to B^+K^-$ rate in a systematic scan over B_s^{**} purities. Right: the distribution of the signal Q-value in the scan. The width of the fitted Gaussian provides the value for the systematic error on the parameters.*

8.5.4 Statistical and Systematic Uncertainties

A number of systematic checks have been performed to determine the systematic error. Using the output of a neural network for signal enrichment in the data makes the exact knowledge of the contribution of single variables to the total error difficult. Nevertheless, to check the stability of the results with respect to the network, a series of network cuts have been applied to the data to achieve different signal purities. The purity range covered was 9% to 26% in 9 steps. The purity range quoted was obtained in Monte Carlo and is defined as the signal divided by the total number of entries in the Q-value distribution and not only at the Q-value of the signal peak.

A further systematic uncertainty arises from the application of the kaon identification network KaonNet. Since the absolute kaon purity in the data is unknown, the uncertainty of the kaon enrichment is a source of systematic error. To check this, three independent B_s^{**} enrichment networks have been trained with different kaon purities chosen at training time. The kaon purities taken were: 42%, 67% and 80%. The systematic checks concerning the B_s^{**} purities were repeated with all three settings.

In order to test the stability of the result as a function of Q-value resolution obtained from the QCorr network defined earlier, the chosen quality was varied between 25%, 50%, 75% and 100% of the best tracks. The widths of the signal were well described by the detector resolution in all four cases. The width measurement shown in figure 8.22 shows a scan for the 50% best resolution class. Accordingly, the fitted widths are smaller for the 25% class and larger for the 75% and 100% classes.

Altogether, the number of fits performed was $9 \times 3 \times 4 = 108$. The results of the fits are presented in figure 8.23 showing the variation of the rate and Q-value measurement as a function of the contributing cuts. Since there was no defined working point, the individual contribution of the variables cannot be given. This is due to the fact that one variable cannot be changed while the others are held fixed. It is only possible to give the total systematic error defined by the width of a Gaussian fitted to the rate and Q-value scan distributions. The statistical error on the measurements has been evaluated by taking the mean of the individual errors on the single points. This method could be applied due to the approximately constant size of the errors over all measurement points.

There have been no checks made on the dependence of the result on the following:

- The modelling of the background. The correction of the background shape has only been performed by a linear function, thus assuming that the deviation between the simulated and true background distributions is only linear. A correction function with more freedom like a third or fourth order polynomial could be used for this test.

- A systematic coming from the fact that the signal shape and width has always been determined from a sample at 80 MeV. Though, this is not expected to add a significant contribution to the systematic error since candidates with a spread of ± 10 MeV around 80 MeV were taken to determine the resolution. This can become problematic if the fit would prefer Q-values which are more than ± 30 MeV further away from the anticipated 80 MeV. The spread of the Q-value measurement is ≈ 4 MeV (see figure 8.23), thus leaving the ignorance of this systematic contribution on the safe side.

8.5.5 Significance of the Measurement

The significance of the rate measurement and consequently the significance of the observation of B_s^{**} mesons is crucial to the importance of the analysis. The quoted result of the rate $\frac{\sigma(B_s^{**})\ \mathrm{Br}(B_s^{**} \to BK)}{\sigma_b} = 0.0085 \pm 0.0023 \pm 0.0013$ suggests a significance of 3.7 standard deviations on the basis of the statistical error assuming Gaussian errors. Two additional tests have been performed to check whether the observed peak is a fluctuation or not. Firstly, assuming the background fit to be correct a check has been performed to obtain the likelihood that the excess in data over the fitted background is a fluctuation or not. Secondly, a check has been performed whether the peak can be absorbed into the background definition by leaving the signal component out of the fit. It is expected to obtain a different significance level from the one quoted above since the errors are not Gaussian, in opposition to the simplified Gaussian assumption above. If the Gaussian assumption

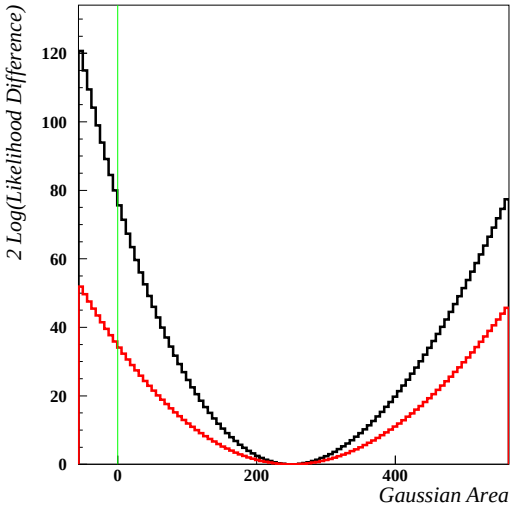

Figure 8.24: *The significance of the B_s^{**} rate measurement.*

were true, the following considerations would provide the same significance result.

The first series of checks makes the assumption that the fitted form of the background is the optimal solution. The significance of the peak is tested by leaving the background as it is and adding to it the signal component with a preset area. For a given signal area, the likelihood that the sum of the two histograms (background+signal) describes the data is calculated. The signal area is varied in the range from -150% to +150% of its nominal value and the likelihood is stored in every step. An example: If the the fit result for the signal area was 100, the value of the signal area is set in the scan to the values between -50 and +250 (in 100 steps). The obtained likelihood function is shown in figure 8.24 (steep function, black) and is defined as $2(\log(L_{scan,i}) - \log(L_{minimum}))$. The y-coordinate at zero area corresponds to the $\Delta\chi^2$ of the null hypothesis: $\Delta\chi^2 = 75.69$. As a conclusion, the observed B_s^{**} signal has a significance of 8.7 standard deviations **not** to come from a fluctuation in the background.

Since the fitted form of the background cannot be assumed to be perfect, a second check is performed including the parameters of the background. This procedure is a new fit with the constraint that the area of the signal is set by hand in a scan series. The signal area is varied between -150% and +150% of its nominal value. Also the mean of the signal component is kept to be the same as in the full fit. The fitted parameters are the normalization and slope correction of the background. The negative log likelihood of the

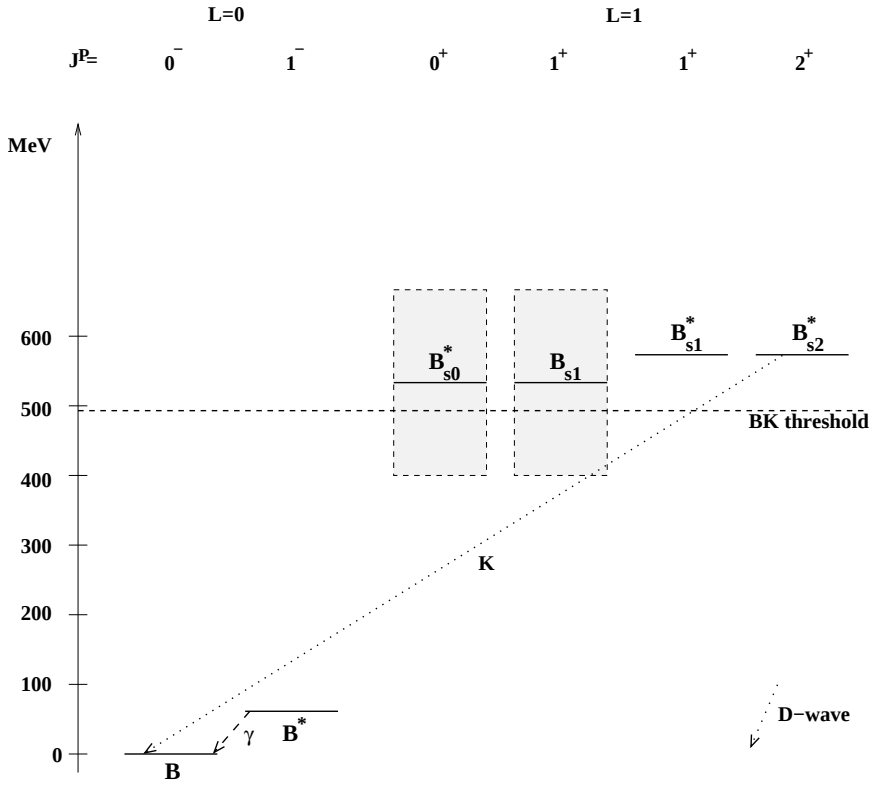

Figure 8.25: *Measured component of the B_s^{**} spectrum.*

full fit result is subtracted off the negative log likelihood of the fit in the scan. To match the definition of the χ^2 distribution, the difference is multiplied by two. The result corresponds to a $\Delta\chi^2$ difference due to the deviation from the optimal parameters of the full fit. Displayed in figure 8.24 (shallow curve, red) is the obtained likelihood difference as a function of the preset signal area. Reading off the y-coordinate of the likelihood function at zero signal area gives for the null hypothesis: $\Delta\chi^2 = 33.64$. As a result, the observed B_s^{**} signal has a significance of 5.8σ **not** to be absorbable into the background. The inclusion of the background shape uncertainty in the calculation reduced the significance of the signal, but is still well above the 3σ level usually taken as a border for an observation. The significance obtained on the basis of the statistical error and assuming Gaussian errors has been proven to be too conservative and to be too low.

8.5.6 Interpretation

The application of the ideas of HQET to B^{**} mesons results in the prediction of four states grouped into two doublets $(0^+, 1^+)$ and $(1^+, 2^+)$. Figure 8.25

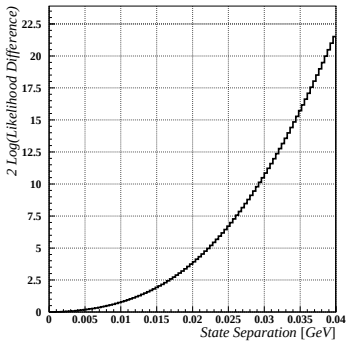

Figure 8.26: *Likelihood difference for the scenario including two states (the 1^+ and 2^+) at 80 MeV contributing to the observed peak. The plot shows the scanned distance between the two states and on the y-axis two times the log likelihood difference to the optimal full fit result.*

shows the predicted masses of the two ground state B mesons and of the four orbitally excited B_s mesons. The doublet corresponding to the total light spin $j_q = 1/2$ is assumed to be broad and the doublet corresponding to the total light spin $j_q = 3/2$ is assumed to be narrow. The decay of the 2^+ can occur via two modes: into B^*K and BK. In total, the anticipated number of observable peaks in the Q-value distributions is thus five: 2 broad and 3 narrow. The fitted data of figure 8.19 is consistent with only one narrow resonance. The remaining four components of the B_s^{**} spectrum remain unobserved. The implications of this observation and the possible interpretation will be discussed in the following.

The width of the broad B_s^{**} states is assumed to be (this is not proven yet) of the order of a hundred MeV. In the fits performed, there is no attempt made to measure these states. In contrast, with the given background definition, the broad states are explicitely included in the background. This is the case due to the normalization and slope correction of the background. Any very broad state located in the Q-value range of the fitted plots can thus be absorbed into the normalization correction. As a consequence, the measurement is fully insensitive to the broad states. The masses of the broad states 0^+ and 1^+ in the spectrum of figure 8.25 can not be measured in this analysis.

As mentioned above and discussed earlier in section 8.5.3, the narrowness of the observed peak excludes more than two states contributing to the resonance if these states are separated in mass. Possible interpretations of the measured peak are:

- The resonance is due to the sum of the 1^+ and 2^+ decay into B^*K, with a low mass splitting between the states (A splitting of 12 MeV is excluded at a 1σ level). As a consequence, the decay $B_{s2}^* \to BK$ would have to show up 46 MeV higher in Q-value. Although a vanishing mass splitting is not ruled out by HQET, measurements in the B^{**} and D^{**} sector give no hint in this direction. The mass splitting between the

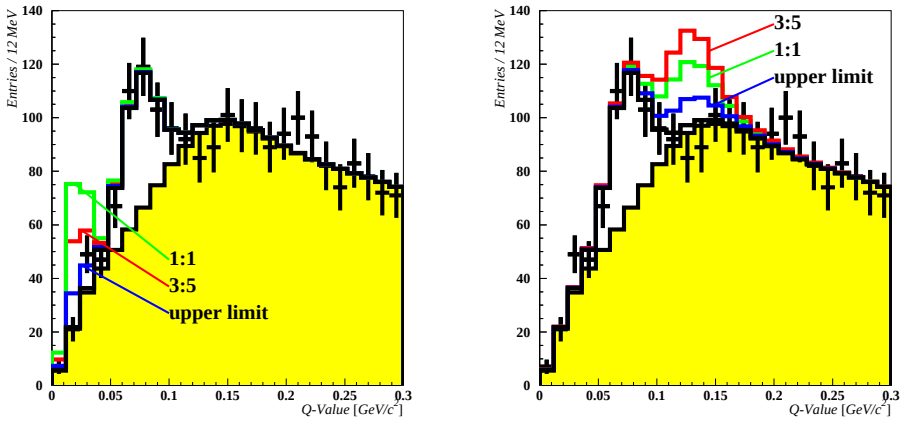

Figure 8.27: *Upper limits and HQET expectations of unseen B_s^{**} components.*

D_{s1} and the D_{s2}^* mesons is approximately 37 MeV. Taking into account the $1/m$ corrections of HQET, the anticipated mass splitting between the B_{s1} and B_{s2}^* states is $37 \cdot \frac{m_c}{m_b} \approx 12$ MeV.

- The resonance is due to the $B_{s2}^* \to BK$ decay channel. The B^*K decay channel of the 1^+ and 2^+ states would then have to show up 46 MeV lower in Q-value.

Since no evidence for either second peak is observed in the data, an upper limit for the production rate and branching ratios of the unseen narrow B_s^{**} states is evaluated. If the measured peak was due to the addition of the 1^+ and 2^+ state decaying into B^*K, the decay of the 2^+ into BK would be approximately 46 MeV higher in Q-value since the B_{s2}^* is allowed to decay into both BK and B^*K. The partial widths scale roughly like $p^{2L+1} = p^5$, thus the production ratio is

$$\frac{\Gamma(B_{s2}^* \to BK)}{\Gamma(B_{s2}^* \to B^*K)} = \left(\frac{354.3}{277.6}\right)^5 = 3.4 \tag{8.11}$$

at the measured Q-value and with p measured in MeV. Taking into account the better acceptance at 126 MeV Q-value (13% better compared to the acceptance at 80 MeV Q-value), this ratio is even more enhanced in the measurement. Assuming a production ratio 5:3 for $B_{2s}^* : B_1$ (spin counting), the expected HQET signal would be the red solid line in figure 8.27 (right-hand side plot). Analogously, if the $B_{2s}^* : B_1$ production ratio was 1:1 (state counting), the HQET expected signal would be the green line. The upper limit at 95% confidence extracted from the data on the signal is

$$\frac{\sigma(B_{s2}^*)\ \mathrm{Br}(B_{s2}^* \to BK)}{\sigma_b} < 0.0014 \quad @\ 95\%\ \mathrm{CL}, \tag{8.12}$$

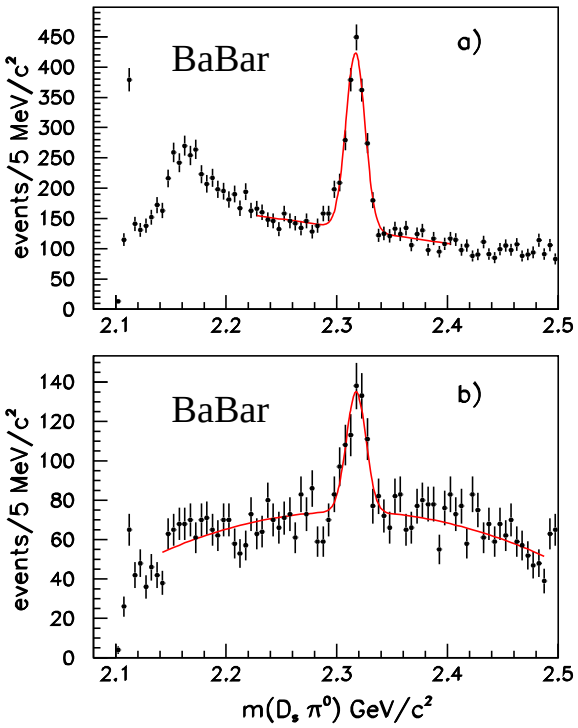

Figure 8.28: *The BaBar $D_s^+\pi^0$ mass distribution for (a) the decay $D_s^+ \to K^+K^-\pi^+$ and (b) the decay $D_s^+ \to K^+K^-\pi^+\pi^0$ [A+03].*

and is shown in blue. The upper limit was based, in accordance with the method suggested by the Particle Data Group [H+02], on the rate measurement -0.00210 ± 0.00123 and using the formula

$$1 - \alpha = \int_{-\infty}^{s_{up}} p(s|n)ds = \frac{\int_{-\infty}^{s_{up}} L(n|s)\pi(s)ds}{\int_{-\infty}^{\infty} L(n|s)\pi(s)ds} \tag{8.13}$$

where α was set to 5% to give the 95% confidence level limit and the prior $\pi(s)$ was set to be zero for $s < 0$, thus reflecting the Bayesian knowledge that the rate measurement can only be positive. As a consequence, the B^*_{s2} would mainly decay into $B * K$ with the branching ratio

$$\frac{\Gamma(B_{s2} \to BK)}{\Gamma(B^*_{s2} \to B * K)} < 32.9\% \quad @\ 95\%\ \text{CL} \tag{8.14}$$

which is very unlikely.

Interpreting the observed peak as the decay $B^*_{s2} \to BK$ implies that 46 MeV lower in Q-value, a sum of the decays B^*_{s2} and B_{s1} into B^*K has to be located. The resonances of the B^*_{s2} and B_{s1} are expected to be separated

by the hyperfine splitting between these states. Due to the lower acceptance ($\approx 50\%$) at 34 MeV Q-value compared to the acceptance at 80 MeV Q-value and due to the p^{2L+1} suppression factor of

$$\left(\frac{p(B_{s2}^* \to BK)}{p(B_{s2}^* \to B^*K)}\right)^{2L+1} = \left(\frac{277.5}{177.7}\right)^5 = \frac{1}{9.3} \tag{8.15}$$

the B_{s2}^* is expected to decay mainly (over 95%) into BK. Assuming again a $B_{2s}^* : B_{1s}$ production ratio between 1:1 and 5:3, the acceptance corrected HQET expectation for the signal is displayed in figure 8.27 (left-hand side plot). Also included in the figure is the extracted confidence limit for the produced peak corresponding to an upper limit on the production rate of

$$\frac{\sigma(B_{s1})\ \mathrm{Br}(B_{s1} \to B^*K)}{\sigma_b} < 0.00148 \quad @\ 95\%\ \mathrm{CL}. \tag{8.16}$$

Comparing the measurement to the HQET expectation, an upper limit for the branching ratio

$$\frac{\Gamma(B_{s1} \to B^*K)}{\Gamma(B_{s2}^* \to BK)} < 39.5\% \quad @\ 95\%\ \mathrm{CL} \tag{8.17}$$

can be given. An alternative decay mode of the B_s^{**} mesons is into the final state $B_s^{(*)}(\pi^+\pi^-)_{s-wave}$, which is not isospin violating and which occurs via a P-wave transition. The decay channel $B_{s2}^* \to B_s(\pi^+\pi^-)_{s-wave}$ is forbidden by parity conservation, thus the alternative decay modes are more enhanced for the B_{s1} meson. In addition, the low decay momentum of the B_{s1} meson ($p = 142$ MeV/c) compared to the decay momentum of the B_{s2}^* meson ($p = 277$ MeV/c) could be responsible for a relatively larger suppression of the kaonic decay channel of the 1^+ state.

Latest results on the spectroscopy of excited states in the D sector provided by the BaBar Collaboration [A$^+$03] offer yet another possible scenario. The observed peak at 2317 MeV mass (figure 8.29) is compatible with a $J^P = 0^+$ assignment. The measurement of BaBar has already been confirmed by CLEO [B$^+$03], having additionally observed a narrow resonance at a mass of 2.46 GeV/c^2 in the $D_s^{*+}\pi^0$ final state. They confirm that the two states are consistent with an interpretation as $c\bar{s}$ mesons with orbital momentum $L = 1$, and spin-parity $J^P = 0^+$ and 1^+. A possible explanation of the measurement is given by Barnes, Close and Lipkin [BCL03] who propose that the resonance is due to a DK bound state **below** the DK threshold. Analogously, the non-observation of the B_{s1} state and similarly of the broad B_s^{**} states could be due to a mechanism shifting the masses of these states below the BK and B^*K thresholds. A similar measurement to that performed by the CLEO and BaBar Collaborations involving the decay channel $B_s^{**} \to B_s^{(*)}\pi^0$ could shed light into the whereabouts of these states.

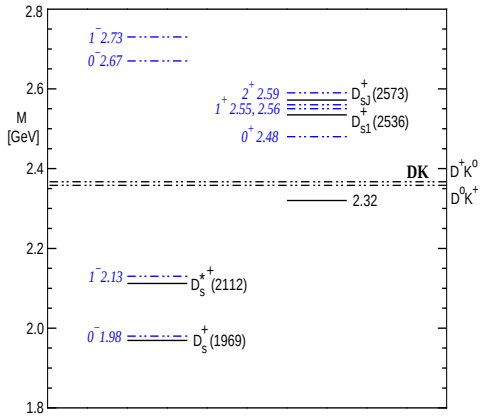

Figure 8.29: *The experimental (solid) and theoretical (dashed) spectrum of $c\bar{s}$ mesons. DK thresholds and the 2.32 BaBar state are also shown [BCL03].*

This analysis is not feasible at DELPHI due to the very low π^0 reconstruction efficiency and the very low B_s statistic. There is hope, that the LHCb Collaboration will be able to measure this channel due to their excellent π^0 reconstruction efficiency over a large π^0 energy range (3 to 50 GeV) [LHC].

In summary, the data show a clear and narrow resonance at 80 MeV Q-value at a constant rate with respect to all considered enrichment and quality cuts. The resonance is compatible with being the 2^+ state of the B_s^{**} mesons decaying into BK. The assignment to the 1^+ state can be experimentally ruled out due to the non-observation of a signal at 46 MeV higher Q-value. The 1^+ state, expected to be approximately 58 MeV lower in Q-value, is not seen, though this may be due to the degraded measurement acceptance and due to the special form of the background in this range of the spectrum. For this state, only an upper limit can be extracted showing a hint towards alternative decay modes rather than the $B^{(*)}K$ decay mode of B_s^{**} mesons and that the coupling to these is a strong function of the B_s^{**} masses. No results have been obtained for the broad states due to the insensitivity of the experimental method to these resonances. Indications from other experiments in the D sector show possible complications on their masses thus making their measurement in the BK channel possibly even impossible.

Chapter 9

Summary

The mass and rate of orbitally excited B mesons was measured using multihadronic Z^0 decays taken in the years 1994 to 1998 by the DELPHI detector. B_s^{**} mesons were analyzed using inclusively reconstructed B mesons combined with kaon candidates. For the reconstruction of the mass, more exactly the Q-value of the decay, a neural network approach has been applied considerably improving the resolution. The Q-value distribution has been enriched in B_s^{**} candidates by a dedicated B_s^{**} network, trained to obtain high purities at still high efficiencies. As a proof of the method, the Q-value reconstruction and the B^{**} enrichment were extended to B_d^{**} and B_u^{**} mesons without claiming a complete analysis of these states.

A narrow $B_{u,d}^{**}$ signal has been observed at a Q-value of

$$Q(B_{u,d}^{**} \to B^{(*)}\pi)_{narrow} \quad = \quad 285 \pm 3(\text{stat}) \pm 7(\text{syst})\text{MeV/c}^2 \qquad (9.1)$$

with a Gaussian width of $\sigma_Q(B_d^{**}) = 49 \pm 5(\text{stat}) \pm 6(\text{syst})\text{MeV/c}^2$ for the B_d^{**} and a width of $\sigma_Q(B_u^{**}) = 60 \pm 5(\text{stat}) \pm 6(\text{syst})\text{MeV/c}^2$ for the B_u^{**}. The signal can be described consistently as originating from the narrow $B_{u,d}^{**}$. The mass of orbitally excited states can be extracted to be

$$m(B_{u,d}^{**} \to B^{(*)}\pi)_{narrow} \quad = \quad 5703 \pm 3(\text{stat}) \pm 7(\text{syst})\text{MeV/c}^2 \ . \qquad (9.2)$$

The production rate of narrow $B_{u,d}^{**}$ is

$$\frac{\sigma(B_{u,d}^{**})_{narrow}\text{Br}(B^{**} \to B^{(*)}\pi)}{\sigma_b} \quad = \quad 0.193 \pm 0.02(\text{stat}) \pm 0.01(\text{syst}), \qquad (9.3)$$

i.e. approximately 20% of each b-jet contains a narrow B_u^{**} or B_d^{**} meson.An additional broad component has been observed in the data at the Q-value

$$Q(\text{Broad}) \quad = \quad 480 \pm 20(\text{stat}) \pm 18(\text{syst})\text{MeV/c}^2 \ , \qquad (9.4)$$

fitted with a Breit-Wigner function. The given Q-value corresponds to the most probable value of the function. The width (FWHM) of the resonance

is

$$\text{FWHM} \quad = \quad 574 \pm 90(\text{stat}) \pm 80(\text{syst})\text{MeV/c}^2 \tag{9.5}$$

and its rate is extracted to be

$$\frac{\sigma(B^{**}_{u,d})_{broad}\text{Br}(B^{**} \to B^{(*)}\pi)}{\sigma_b} = 0.16 \pm 0.04(\text{stat}) \pm 0.04(\text{syst}) \ . \tag{9.6}$$

Its calculated rate is dependent on the broad to narrow ratio R also fitted from the data. Its value is

$$R \quad = \quad 0.8 \pm 0.2(\text{stat}) \pm 0.2(\text{syst}) \tag{9.7}$$

which is compatible with the predictions of HQET lying between 1:2 and 1:1. In conclusion, the results are in general agreement with previous measurements for the narrow states. There is evidence for the existence of a broad resonance at high Q-values possibly coming from broad $B^{**}_{u,d}$ or from radially excited B mesons. The stability of the signals has been checked in systematic studies showing a consistent result.

The major aim of the thesis, the analysis of excited B_s mesons has been performed in the decay $B^{**}_s \to B^{(*)\pm}K^{\mp}$. The extensive use of neural network techniques, especially for kaon identification, has made it possible to establish, after a first observation by the OPAL experiment, the existence of B^{**}_s mesons. From the predicted four states one can be seen in the data. Its Q-value is

$$Q(B^{**}_s \to BK) \quad = \quad 80.2 \pm 4.7(\text{stat}) \pm 4.2(\text{syst})\text{MeV/c}^2 \tag{9.8}$$

corresponding to a mass of

$$Q(B^{**}_s) = 5853.2 \pm 4.7(\text{stat}) \pm 4.2(\text{syst})\text{MeV/c}^2 \ , \tag{9.9}$$

with a width of approximately $\sigma_Q(B^{**}_s) \approx 16$ MeV, compatible with detector resolution. The rate of the fitted excess is

$$\frac{\sigma(B^{**}_s)\ \text{Br}(B^{**}_s \to BK)}{\sigma_b} \quad = \quad 0.0085 \pm 0.0023(\text{stat}) \pm 0.0013(\text{syst}) \ , \tag{9.10}$$

and is compatible to the narrow 2^+ state decaying via BK and is compatible with the HQET expectation of 0.0125. The unseen peaks are the broad states and the narrow 1^+ state. The analysis is not sensitive to the broad states due to their possible inclusion into the background definition. This is caused by the uncertainty of the modelling of the background. The background is fitted to the data by a normalization and slope correction factor which can easily absorb any very broad resonance. An upper limit can be

given for the 1^+ narrow resonance. Accepting the seen peak to belong to the 2^+ state, the narrow B_{s1} rate is

$$\frac{\sigma(B_{s1})\ \mathrm{Br}(B_{s1} \to B^*K)}{\sigma_b} < 0.00148 \quad @\ 95\%\ \mathrm{CL}. \tag{9.11}$$

Comparing the measurement to the HQET expectation, which is about three times higher, an upper limit for the branching ratio

$$\frac{\Gamma(B_{s1} \to B^*K)}{\Gamma(B^*_{s2} \to BK)} < 39.5\% \quad @\ 95\%\ \mathrm{CL} \tag{9.12}$$

can be given. Further analyses, e.g. at CDF or LHCb, are needed to confirm the result for the B^*_{s2} and to establish the mass of the unseen states.

List of Figures

List of Tables

Bibliography

[A+86] H. Albrecht et al., Phys. Rev. Lett. **56** (1986), 549.

[A+89a] H. Albrecht et al., Phys. Lett. **B 221** (1989), 422.

[A+89b] H. Albrecht et al., Phys. Lett. **B 231** (1989), 208.

[A+95a] P. Abreu et al., *B^* production in Z decays*, Z. Phys. **C68** (1995), 353–362.

[A+95b] P. Abreu et al., *Observation of orbitally excited B mesons*, Phys. Lett. **B345** (1995), 598–608.

[A+95c] R. Akers et al., *Observations of $\pi - B$ charge - flavor correlations and resonant $B\pi$ and BK production*, Z. Phys. **C66** (1995), 19–30.

[A+96a] P. Abreu et al., *Tuning and test of fragmentation models based on identified particles and precision event shape data*, Z. Phys. **C73** (1996), 11–60.

[A+96b] H. Albrecht et al., *Measurement of the decay $D_{s2}^{*+} \to D^0 K^+$*, Z. Phys. **C69** (1996), 405–408.

[A+98] P. Abreu et al., *First Evidence for a Charm Radial Excitation, D*'*, Phys. Lett. **B426** (1998), 231–242.

[A+99] M. Acciarri et al., *Measurement of the spectroscopy of orbitally excited B mesons at LEP*, Phys. Lett. **B465** (1999), 323–334, hep-ex/9909018.

[A+00a] P. Abreu et al., *Inclusive Σ- and $\Lambda(1520)$ production in hadronic Z decays*, Phys. Lett. **B475** (2000), 429–447, hep-ex/0103020.

[A+00b] S. Anderson et al., *Observation of a broad $L = 1$ $c\bar{q}$ state in $B^- \to D^{*+}\pi^- pi^-$ at CLEO*, Nucl. Phys. **A663** (2000), 647–650, hep-ex/9908009.

[A+01a] G. Abbiendi et al., *A Search for a Radial Excitation of the $D^{*\pm}$ Meson*, Eur. Phys. J. **C20** (2001), 445–454, hep-ex/0101045.

[A+01b] P. Abreu et al., *Measurement of $V(cb)$ from the decay process $\bar{B}^0 \to D^{*+} l^- \bar{\nu}$*, Phys. Lett. **B510** (2001), 55–74, hep-ex/0104026.

[A+01c] T. Affolder et al., *Observation of orbitally excited B mesons in $p\bar{p}$ collisions at $\sqrt{s}$=1.8 TeV*, Phys. Rev. **D64** (2001), 072002.

[A+02] G. Abbiendi et al., *Investigation of the decay of orbitally-excited B mesons and first measurement of the branching ratio $BR(B_J^* \to B^*\pi(X))$*, Eur. Phys. J. **C23** (2002), 437–454, hep-ex/0010031.

[A+03] B. Aubert et al., *Observation of a narrow meson decaying to $D_s^+\pi^0$ at a mass of 2.32 GeV/c^2*, hep-ex/0304021.

[AAB+01] Z. Albrecht, T. Allmendinger, G. Barker, M. Feindt, C. Haag, and M. Moch, *BSAURUS: A package for inclusive B reconstruction in DELPHI*, hep-ex/0102001.

[AFM99] Zoltan Albrecht, Michael Feindt, and Markus Moch, *MACRIB: High efficiency - high purity hadron identification for DELPHI*, hep-ex/0111081.

[AGIS83] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983), 31.

[AK+00] A. Ali Khan et al., *Heavy-light mesons and baryons with b quarks*, Phys. Rev. **D62** (2000), 054505, hep-lat/9912034.

[And97] Bo Andersson, *The Lund model*, Cambridge Monogr. Part. Phys. Nucl. Phys. Cosmol. **7** (1997), 1–471.

[B+94] T. Bergfeld et al., Phys. Lett. **B 340** (1994), 194.

[B+96] D. Buskulic et al., *Production of excited beauty states in Z decays*, Z. Phys. **C69** (1996), 393–404.

[B+97] D. Buskulic et al., *Measurements of $|V(cb)|$, form factors and branching fractions in the decays $\bar{B}^0 \to D^{*+} l^- \bar{\nu}_l$ and $\bar{B}^0 \to D^+ l^- \bar{\nu}_l$*, Phys. Lett. **B395** (1997), 373–387.

[B+98] R. Barate et al., *Resonant structure and flavour tagging in the $B\pi^\pm$ system using fully reconstructed B decays*, Phys. Lett. **B425** (1998), 215–226.

[B+03] D. Besson et al., *Observation of a narrow resonance of mass 2.46 GeV/c^2 in the $D_s^{*+}\pi^0$ final state, and confirmation of the $D_{sJ}^*(2317)$*, hep-ex/0305017.

[BCL03] T. Barnes, F. E. Close, and H. J. Lipkin, *Implications of a DK molecule at 2.32 GeV*, hep-ph/0305025.

[Bis95] C. Bishop, *Neural Networks for Pattern Recognition.*, Oxford University Press, 1995.

[BLMÅ02] L. Berntzon, L. Leinonen, T. Moa, and B. Åsman, *Production of Ξ_c^0 and Ξ_b in Z^0 decays in the DELPHI detector at LEP*, DELPHI Note (2002), no. 078, CONF 612.

[BM95] G. Borisov and C. Mariotti, *Fine tuning of track impact parameter resolution of the DELPHI detector*, DELPHI Note (1995), no. 142, PHYS 567.

[BM96] G. Borisov and C. Mariotti, *Fine tuning of track impact parameter resolution of the DELPHI detector*, Nucl. Instrum. Meth. **A372** (1996), 181–187.

[Boua] C. Bourdarios, *Hadron Identification with the DELPHI Ring Imaging Cerenkov Detector*, Prepared for 1994 Meeting of the American Physical Society, Division of Particles and Fields (DPF 94), Albuquerque, New Mexico, 2-6 Aug 1994.

[Boub] C. Bourdarios, *User Guide of the Hadident Rich software.*

[Bra92] S. Brandt, *Datenanalyse.*, BI-Wissenschaftsverlag (Mannheim), 1992.

[Cab63] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963), 531–532.

[Car92] A. Carling, *Introducing Neural Networks.*, Sigma Press, 1992.

[CMR+89] J. Cuevas, J. Marco, A. Ruiz, F. Richard, and F. Simonetto, *Fast Simulation of the Delphi Detector*, Nucl. Instrum. Meth. **A274** (1989), 459–468.

[Del89a] *DELPHI Data Analysis Program (DELANA) User's Guide*, DELPHI Note (1989), no. 44, PROG 137.

[Del89b] *DELPHI Event Generation and Detector Simulation - Reference Manual*, 1989.

[DEL95] DELPHI Collaboration, *Performance of the DELPHI Detector*, 1995, CERN-PPE/95-194.

[DPE01] Massimo Di Pierro and Estia Eichten, *Excited heavy-light systems and hadronic transitions*, Phys. Rev. **D64** (2001), 114004, hep-ph/0104208.

[EFG97] D. Ebert, R. N. Faustov, and V. O. Galkin, *Exclusive nonleptonic decays of B mesons*, Phys. Rev. **D56** (1997), 312–320, hep-ph/9701218.

[EGF98] D. Ebert, V. O. Galkin, and R. N. Faustov, *Mass spectrum of orbitally and radially excited heavy-light mesons in the relativistic quark model*, Phys. Rev. **D57** (1998), 5663–5669, hep-ph/9712318.

[EHQ93] Estia J. Eichten, Christopher T. Hill, and Chris Quigg, *Properties of orbitally excited heavy - light mesons*, Phys. Rev. Lett. **71** (1993), 4116–4119, hep-ph/9308337.

[EHQ94] E. Eichten, C. Hill, and C. Quigg, *Spectra of Heavy-Light Mesons*, FERMILAB (1994), CONF 94.

[Fau94] L. Fausett, *Fundamentals of Neural Networks.*, New York: Prentice Hall, 1994.

[FGM95] R. N. Faustov, V. O. Galkin, and A. Yu. Mishurov, *Exclusive semileptonic decays of B mesons into light mesons in the relativistic quark model*, Phys. Lett. **B356** (1995), 516–524, hep-ph/9505321.

[FM96] Adam F. Falk and Thomas Mehen, *Excited heavy mesons beyond leading order in the heavy quark expansion*, Phys. Rev. **D53** (1996), 231–240, hep-ph/9507311.

[GEF02] V. O. Galkin, D. Ebert, and R. N. Faustov, *Heavy quark potential and mass spectra of heavy mesons*, AIP Conf. Proc. **619** (2002), 336–345, hep-ph/0110190.

[GJ95] Suraj N. Gupta and James M. Johnson, *Quantum chromodynamic potential model for light heavy quarkonia and the heavy quark effective theory*, Phys. Rev. **D51** (1995), 168–175, hep-ph/9409432.

[GK91] Stephen Godfrey and Richard Kokoski, *The Properties of P Wave Mesons with one Heavy Quark*, Phys. Rev. **D43** (1991), 1679–1687.

[GMaL02] Barker G., Feindt M., and Kerzel U. andRamler L., *A Study of the b-Quark Fragmentation Function with the DELPHI Detector at LEP I*, DELPHI 2002-069 CONF 603.

[GNR93] Michael Gronau, Alex Nippe, and Jonathan L. Rosner, *Method for flavor tagging in neutral B meson decays*, Phys. Rev. **D47** (1993), 1988–1993, hep-ph/9211311.

[GR94] Michael Gronau and Jonathan L. Rosner, *Identification of Neutral B Mesons using Correlated Hadrons*, Phys. Rev. **D 49** (1994), 254, hep-ph/9308371.

[H+02] K. Hagiwara et al., *Review of particle physics*, Phys. Rev. **D66** (2002), 010001.

[Hay94] S. Haykin, *Neural Networks: A Comprehensive Foundation.*, Macmillan Publishing, 1994.

[Isg98] Nathan Isgur, *Spin-orbit inversion of excited heavy quark mesons*, Phys. Rev. **D57** (1998), 4041–4053.

[KFP96] C. Kreuter, M. Feindt, and O. Podobrin, *Elephant Reference Manual*, 1996.

[Kol57] A.N. Kolmogorov, *On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition.*, Doklady Akademiia Nauk, 1957.

[LF+90] J. Lee-Franzini et al., *Hyperfine splitting of B mesons and B_s production at the $\Upsilon(5S)$*, Phys. Rev. Lett. **65** (1990), 2947–2950.

[LHC] *LHCb: Calorimeters Technical Design Report*, CERN-LHCC-2002-036.

[Lib] CERN Program Library, URL: http://wwwinfo.cern.ch/asdoc/-cernlib.html.

[Lon92] Leif Lonnblad, *ARIADNE version 4: A Program for simulation of QCD cascades implementing the color dipole model*, Comput. Phys. Commun. **71** (1992), 15.

[MF96] C. Weiser M. Feindt, W. Oberschulte, *How to Use the MAMMOTH Program*, 1996.

[Mor89] Duncan A. Morris, *Heavy Quark Fragmentation Functions in a Simple String Model*, Nucl. Phys. **B313** (1989), 634.

[MUW02] S. Moch, P. Uwer, and S. Weinzierl, *Scattering amplitudes for $e^+e^- \to 3$ jets at next-to-next-to-leading order QCD*, hep-ph/0210009.

[MW00] Aneesh V. Manohar and Mark B. Wise, *Heavy quark physics*, Cambridge Monogr. Part. Phys. Nucl. Phys. Cosmol. **10** (2000), 1–191.

[Neu94] Matthias Neubert, *Theoretical update on the model independent determination of $|V(cb)|$ using heavy quark symmetry*, Phys. Lett. **B338** (1994), 84–91, hep-ph/9408290.

[OG01] LEP B Oscillation and Lifetime Working Group, *http://lepbosc.web.cern.ch/LEPBOSC*.

[Pat96] D. Patterson, *Artificial Neural Networks.*, Singapore: Prentice Hall, 1996.

[PSSZ83] C. Peterson, D. Schlatter, I. Schmitt, and Peter M. Zerwas, *Scaling Violations in Inclusive e^+e^- Annihilation Spectra*, Phys. Rev. **D27** (1983), 105.

[PT] Phi-T, Physics Information Technologies, URL: http://www.phi-t.de.

[Rip96] B.D. Ripley, *Pattern Recognition and Neural Networks.*, Cambridge University Press, 1996.

[Rod98] Jorge L. Rodriguez, *Hadronic decays of beauty and charm from CLEO*, hep-ex/9901008.

[Sac94] Y. Sacquin, *Description of the DELPHI DST Content*, DELPHI Note **161** (1994), no. PROG 210.

[Sjo] T. Sjostrand, *The PYTHIA and JETSET programs*, Prepared for Particles and Fields 92: 7th Meeting of the Division of Particles Fields of the APS (DPF 92), Batavia, Illinois, 10-14 Nov 1992.

[Sjo94] Torbjorn Sjostrand, *High-energy physics event generation with PYTHIA 5.7 and JETSET 7.4*, Comput. Phys. Commun. **82** (1994), 74–90.

[Sta] StatSoft, *Statistics Textbook*, WWW page, URL: http://www.-statsoft.com/textbook/stathome.html.

[ZVOR95] J. Zeng, J. W. Van Orden, and W. Roberts, *Heavy mesons in a relativistic model*, Phys. Rev. **D52** (1995), 5229–5241, hep-ph/9412269.

Acknowledgements

This work would not have been possible without the contribution of a number of people whom I would like to thank at this point:

- I thank Prof. Dr. Michael Feindt for the opportunity to join the DELPHI experiment and to be able to perform the analysis at the Institute for Experimental Nuclear Physics.
- I thank again Prof. Dr. Michael Feindt and Prof. Dr. Günther Quast for the supervision and co-supervision of my work.
- In this sense, I also thank the whole Institute for providing me with the essential research resources and for the stimulating and friendly atmosphere.
- The financial support of my work was covered by the Land Baden-Württemberg and by the German Federal Research Ministry. I thank them for granting me a scholarship for the whole of my thesis.
- The manuscript has been carefully read by Gary Barker, Uli Kerzel and Markus Moch. They are the ones with whom I worked in the last years very closely and without them, the work would have been dull and impossible to manage.
- Last but not least I would like to thank my friends and family.