Universität Karlsruhe (TH), Institut für Experimentelle Kernphysik (EKP)
IEKP-KA/2009-26

Diplomarbeit

Full Reconstruction and Y(4140) Search at Belle

Daniel Zander

20. Januar 2010

Referent: Prof. Dr. Michael Feindt
Korreferent: Prof. Dr. Günter Quast

Zusammenfassung

Die theoretische Grundlage der Teilchenphysik bildet heutzutage das *Standardmodell der Teilchenphysik*. Sowohl die Überprüfung dieses Modells und seiner möglichen Erweiterungen, als auch die möglichst präzise Bestimmung seiner Parameter sind die Aufgabe der Experimentalphysiker. Im Rahmen dieser Arbeit wird ein sowohl ein mögliches, neues Phänomen, die $Y(4140)$-Resonanz, untersucht, als auch die Grundlage für ein bestimmtes Analysewerkzeug, die vollständige Rekonstruktion, gelegt.

Das Standardmodell der Teilchenphysik

Die moderne Teilchenphysik beschäftigt sich mit den kleinsten Bestandteilen der Materie und den Kräften, die zwischen diesen Teilchen wirken. Dabei werden Objekte untersucht, deren Eigenschaften sich den Erklärungsversuchen der klassischen Physik entziehen. Eine Beschreibung dieser Objekte kann die Quantenfeldtheorie liefern. Sie fußt auf den Gesetzen der Quantenmechanik, die Anfang des 20. Jahrhunderts erstmals formuliert wurden, und stellt das theoretische Grundgerüst für die Teilchenphysik dar. Diese Theorie beschreibt alle bekannten Elementarteilchen und deren Wechselwirkungen und wird deswegen als das Standardmodell der Teilchenphysik bezeichnet.

Fermionen und Bosonen

Das Standardmodell postuliert zwei Arten von Teilchen: Fermionen und Bosonen.
Fermionen sind Teilchen mit halbzahligem Spin[1], die sich wiederum in Leptonen und Quarks unterteilen lassen. Von zunächst nur zwei verschiedenen Quarks (up und down) und einem Paar Leptonen (Elektron und Neutrino) ausgehend, wurde das Standardmodell bis heute auf sechs Quarks und drei Leptonpaare und deren jeweilige Antiteilchen erweitert, die in drei sogenannten Familien gruppiert werden können (siehe Tabelle 0.1). Die geladenen Leptonen $e^{\pm}$, $\mu^{\pm}$ und $\tau^{\pm}$ unterscheiden sich in ihrer Masse und der einer Quantenzahl, dem sogenannten Leptonflavour. sie alle tragen die Elementarladung $\pm 1e$. Die jeweils zu den geladenen Leptonen zugehörigen Neutrinos ν_e, ν_μ und ν_τ wurden als masselose, ungeladene Teilchen postuliert, um die Energie- und Impulserhaltung bei bestimmten Reaktionen zu gewährleisten. Ihre Existenz wurde bald auch experimentell bestätigt.
Die Namen der Quarks, auch Flavour genannt, sind historisch gegeben und haben keine tiefere Bedeutung. Der Flavour ist meist eine Erhaltungsgröße und kann nur in bestimmten Reaktionen verändert werden. Die Quarks tragen ein oder zwei Drittel der

[1] Der Spin ist eine quantenmechanische Eigenschaft von Teilchen und kann mit dem Formalismus des Drehimpulses beschrieben werden.

Familie	Name	Ladung	Masse
Quarks			
I	up (u)	$+\frac{2}{3}$	$2.55^{+0.75}_{1.05}$ MeV
	down (d)	$-\frac{1}{3}$	$5.04^{+0.96}_{-1.54}$ MeV
II	charm (c)	$+\frac{2}{3}$	$1.27^{+0.07}_{-0.11}$ MeV
	strange (s)	$-\frac{1}{3}$	105^{+25}_{-35} MeV
III	top (t)	$+\frac{2}{3}$	$171.3 \pm 1.3 \pm 1.2$ GeV
	bottom (b)	$-\frac{1}{3}$	$4.20^{+0.17}_{-0.07}$ GeV
Leptonen			
I	e	-1	0.511 MeV
	ν_e	0	< 2 eV
II	μ	-1	105.7 MeV
	ν_μ	0	< 2 eV
III	τ	-1	1776.8 MeV
	ν_τ	0	< 2 eV

Tabelle 0.1: Übersicht über die Teilchen im Standardmodell [eaPDGon].

Elementarladung und treten nur in gebundenen Zuständen von zwei oder mehr Quarks auf. In diesem Zusammenhang wird vom Quark-Confinement gesprochen. Zustände mit zwei Quarks (einem Quark und einem Antiquark) werden "Mesonen" genannt, Zustände mit drei Quarks "Baryonen". Im Standardmodell wird den Quarks eine Quantenzahl zugeordnet, die diesen Umstand formal fassbar macht: die Farbe. Konvention ist die Unterscheidung in rot, grün und blau, mit den jeweiligen Antifarben. Mit der Forderung, es dürfe nur farbneutrale Zustände geben, kann das Auftreten von Mesonen (Farbe + Antifarbe) und Baryonen (rot + grün + blau) beschrieben werden. Die theoretisch ebenfalls möglichen gebundenen Zustände von vier, fünf oder mehr Quarks wurden bisher nicht eindeutig beobachtet.
Bosonen sind Teilchen mit ganzzahligem Spin. Zu dieser Kategorie werden unter anderem diejenigen Teilchen gezählt, die die Wechselwirkungen zwischen den Quarks und Leptonen vermitteln. Die Vorstellung, eine Wechselwirkung werde durch den Austausch von Teilchen vermittelt, lässt mit zwei Personen auf rutschigem Untergrund vergleichen, die sich einen schweren Gegenstand zuwerfen und sich dadurch gegenseitig abstoßen.

Wechselwirkungen

Es sind vier verschiedene in der Natur auftretende Wechselwirkungen bekannt. Drei davon werden innerhalb des Standardmodells erklärt: die starke, die schwache und die elektromagnetische Wechselwirkung. Jede dieser Wechselwirkungen kann durch den Austausch spezifischer Bosonen beschrieben werden. Bisher gelang es nicht, eine konsistente Feldtheorie der vierten bekannten Wechselwirkung, der Schwerkraft zu entwickeln. Folglich kann sie in der Teilchenphysik nicht berücksichtigt werden. Die Effekte der Schwerkraft sind verglichen mit den anderen Wechselwirkungen jedoch außerordentlich gering,

sodass sie für sämtliche Reaktionen auf dem Gebiet der Teilchenphysik vernachlässigt werden können. Eine Übersicht über die verschiedenen Austauschbosonen gibt Tabelle 0.2.

Wechselwirkung	Name	Ladung	Masse
elektromagnetisch	Photon (γ)	0	0
schwach	W^+	$+1$	80.403 ± 0.029 GeV
	W^-	-1	
	Z^0	0	91.1876 ± 0.0021 GeV
stark	Gluon g	0	0

Tabelle 0.2: Übersicht über die Austauschbosonen des Standardmodells [eaPDGon].

Die elektromagnetische Wechselwirkung

Alle elektrisch geladenen Teilchen nehmen an der elektromagnetischen Wechselwirkung teil. Dies sind alle Quarks und die geladenen Leptonen. Die Wechselwirkung wird durch den Austausch von masselosen Photonen γ beschrieben. Da die Austauschteilchen masselos sind, ist die Reichweite der elektromagnetischen Wechselwirkung unendlich. Die relative Stärke einer Wechselwirkung wird durch die sogenannte Kopplungskonstante festgelegt. Bei der elektromagnetischen Wechselwirkung heißt die Kopplungskonstante Feinstrukturkonstante α_S und sie kann geschrieben werden als

$$\alpha = \frac{e^2}{4\pi\hbar c} \approx \frac{1}{137} \ . \tag{0.1}$$

Die starke Wechselwirkung

An der starken Wechselwirkung nehmen alle farbetragenden Teilchen, also die Quarks, teil. Sie wird vermittelt durch masselose, selbst Farbe tragende Gluonen, von denen es acht verschiedene Arten gibt. Da die Gluonen selbst Farbe tragen, koppeln sie auch untereinander, was die nur sehr geringe Reichweite der starken Wechselwirkung zur Folge hat. Die starke Kopplungskonstante α_S ist etwa 100 mal größer als die Feinstrukturkonstante:

$$\frac{\alpha_\mathrm{S}}{\alpha} \approx 100 \qquad \Rightarrow \qquad \alpha_\mathrm{S} \approx 1 \ . \tag{0.2}$$

Die Theorie der starken Wechselwirkung kann ebenso erklären, wie es zum Quark-Confinement kommt. Das starke Potential zwischen zwei Quarks sinkt bei geringer Entfernung stark ab, in großer Nähe kann man die Quarks eines gebundenen Zustandes als quasi frei betrachten. Jedoch steigt das Potential mit der Entfernung stark an. Zur räumlichen Trennung von zwei Quarks ist also eine enorme Energiemenge vonnöten, die schon bei geringer Entfernung dazu ausreicht, ein neues Quark-Antiquark-Paar zu bilden. Je ein Quark dieses neuen Paares bildet daraufhin mit einem der beiden ursprünglichen Quarks ein neues Meson. Bei entsprechend hohen Energien kann sich dieser Vorgang häufig wiederholen, was man als Hadronisierung bezeichnet.

Die schwache Wechselwirkung

Alle bekannten Teilchen nehmen an der schwachen Wechselwirkung teil, jedoch sind deren Effekte bei niedrigen Energien relativ gering, sodass sie nur in Erscheinung tritt, wenn die starke und elektromagnetische Wechselwirkung unterdrückt sind. Die schwache Wechselwirkung wird durch massebehaftete Bosonen vermittelt, wodurch sie eine geringe Reichweite hat. Diese Bosonen sind zum Einen das neutrale Z^0-Boson, zum anderen die geladenen $W^{\pm}$-Bosonen. Die schwache Wechselwirkung ist die einzige Wechselwirkung, die den Flavour zu ändern vermag. Das historisch erste Beispiel dafür ist der β-Zerfall des Neutrons, der in Abbildung 0.1 skizziert ist. Aufgrund der hohen Masse der $W^{\pm}$ und

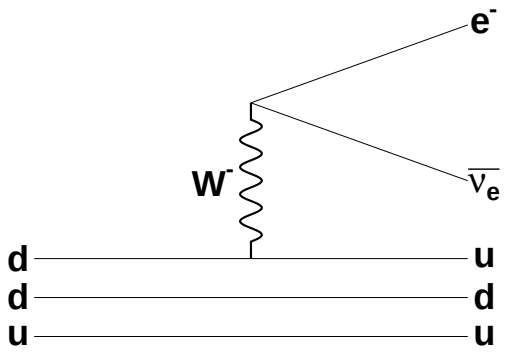


Abbildung 0.1: β-Zerfall des Neutrons.

Z^0 Bosonen ist die schwache Wechselwirkung kurzreichweitig. Die Kopplungskonstante ist mit

$$\alpha_W \sim 10^{-6} \quad \text{für} \quad Q^2 \to 0 \tag{0.3}$$

sehr klein.
Um das eben skizzierte Standardmodell der Teilchenphysik nun zu testen oder seine Parameter genauer zu bestimmen, werden Analysen durchgeführt. Im Rahmen vieler solchen Analysen werden große Datenmengen, die von Detektoren an Teilchenbeschleunigern aufgezeichnet wurden, aufbereitet und statistisch ausgewertet.

Statistische Analyseverfahren

Schnittbasierte Analysen

Während einer typischen Analyse im Rahmen der Teilchenphysik, ist es häufig vonnöten, eine bestimmte Art von Teilchen mit bestimmten Eigenschaften auszuwählen. Beispielsweise sind sogenannte B Mesonen auszuwählen, die auf eine ganz bestimmte Art zerfallen. Es ist also nötig, innerhalb einer Liste von *Kandidaten* das gewünschte *Signal* vom unerwünschten *Untergrund* zu trennen. Diese Trennung kann sehr unterschiedlich

vonstatten gehen.
Die einfachste Methode, Signal von Untergrund zu trennen, also eine Selektion durchzuführen, sind Schnitte auf einzelne, physikalische Variablen. Beispielsweise würde man fordern, dass ein bestimmtes Teilchen eine Masse, einen Winkel zum Strahlrohr oder einen Impuls innerhalb eines engen Bereichs haben soll. Solche Schnitte haben den Vorteil, dass sie besonders einfach durchzuführen sind, jedoch können komplexere Zusammenhänge, in Form von Korrelationen zwischen den Variablen, nicht berücksichtigt werden. Eine schnittbasierte Selektion ist in Abbildung 0.2 visualisiert. Angenommen, der

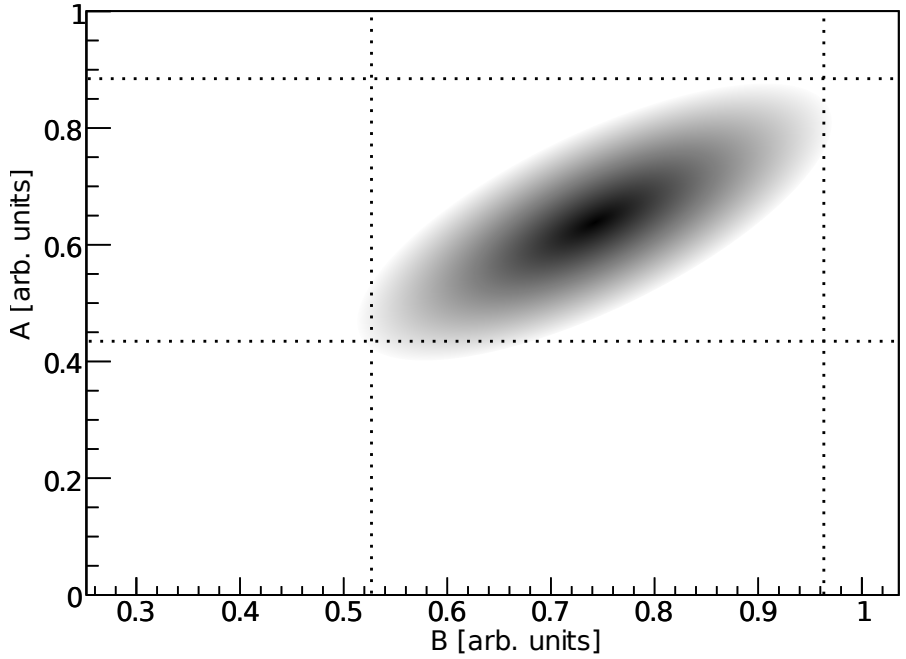


Abbildung 0.2: Visualisierung von Schnitten (gepunktete Linien), schwarz: Signal.

schwarze Bereich stelle das Signal dar und der Untergrund sei über den ganzen Bereich gleichverteilt, so wird deutlich, dass einfache Schnitte ohne Berücksichtigung von Korrelationen keine optimalen Ergebnisse liefern können.

Multivariate Verfahren

Eine Verbesserung kann mithilfe von multivariaten Verfahren erreicht werden, mit denen es möglich ist, Korrelationen zwischen den verschiedenen Variablen zu berücksichtigen. Viele dieser Verfahren sind jedoch in ihrer Ausführung relativ kompliziert und zeitaufwändig. Im Rahmen dieser Arbeit wurde das Softwarepaket NeuroBayes zur Selektion verwendet, was ein einfach zu handhabendes, multivariates Analysewerkzeug ist, das sehr gute Ergebnisse liefert.

NeuroBayes

Das Softwarepakte NeuroBayes besteht im Groben aus zwei Bereichen: Dem Preprocessing und dem neuronalen Netz. Das Preprocessing bereitet die Eingabedaten, wie zum Beispiel Masse-, Winkel- und Impulsverteilungen, aber auch kompliziertere, kombinierte Variablen, wie Likelihood-Ratios, auf eine bestimmte Weise auf. Diese aufbereiteten Variablen werden danach benutzt, um ein neuronales Netz zu trainieren. Ein neuronales Netz kombiniert die einzelnen Eingabevariablen mit bestimmten Gewichtungen versehen miteinander, um für jeden Kandidaten einzeln zu entscheiden, wie signal- bzw. untergrundartig dieser Kandidat ist. Um ein Training durchzuführen ist es vonnöten, einen Satz Kandidaten zur Verfügung zu stellen, bei dem die Wahrheit, also ob ein Kandidat Signal oder Untergrund ist, bekannt ist. Diese sogenannten Monte Carlo Daten werden üblicherweise aus komplexen Simulationen gewonnen. Während des Trainings werden die Gewichtungen, nach denen die Eingabedaten kombiniert werden, derartig angepasst, dass sowohl Signal als auch Untergrund möglichst korrekt von der Ausgabe des neuronalen Netzes beurteilt werden. Vereinfacht ausgedrückt ist das Training eines neuronalen Netzes eine bestimmte Art der mehrdimensionalen Anpassungsrechnung.
Als Ergebnis eines solchen Trainings wird ein sogenannter Experte erstellt, der für jeden Kandidaten die Signal- bzw. Untergrundwahrscheinlichkeit ausgeben kann. Diese einzelne Variable wird schließlich für den Selektionsprozess herangezogen.

Experimenteller Aufbau

Die Analysen in dieser Arbeit wurden mit Daten durchgeführt, die mit dem Belle Detektor am KEK B Beschleuniger in Japan aufgezeichnet wurden. Der KEK B Beschleuniger ist ein asymmetrischer Elektron-Positron-Speicherring. Das bedeutet, dass die Elektronen auf einer höhere Energie ($E_- = 8.0$ GeV) beschleunigt werden als die Positronen ($E_+ = 3.5$ GeV). Dies hat zur Folge, dass das Schwerpunktsystem der Kollisionen gegenüber dem Laborsystem bewegt ist, was für bestimmte Messungen von großem Vorteil ist. Die Schwerpunktsenergie des KEK B Beschleunigers beträgt

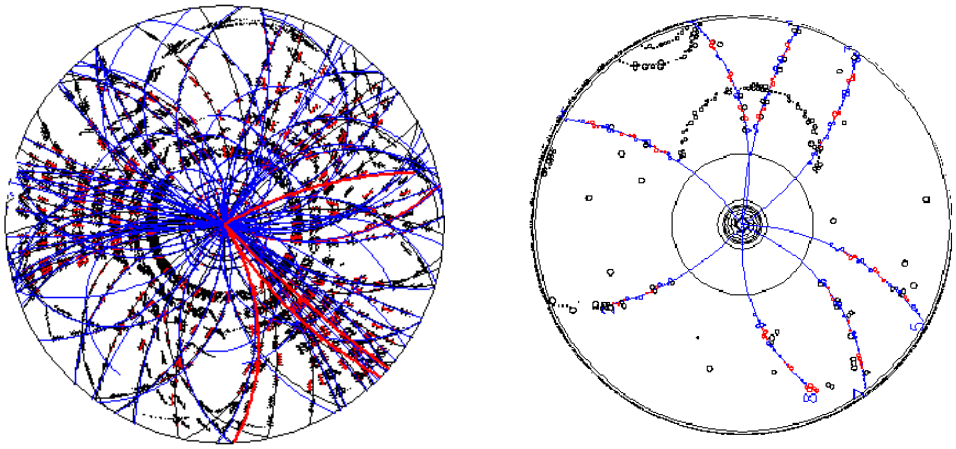
$$\sqrt{s} = \sqrt{4E_+E_-} \approx 10.58 \text{ GeV} . \tag{0.4}$$

Sie liegt damit genau auf der $\Upsilon(4S)$-Resonanz. Diese Resonanz zerfällt nahezu in 100% der Fälle in ein B-Meson-Paar. Die beiden B-Mesonen zerfallen nach kurzer Lebensdauer auf vielfältige Weise. Der Belle-Detektor zeichnet diese Zerfälle mit hoher Genauigkeit auf. Er ist innerhalb eines 1.5 T starken Magnetfeldes, das von einem supraleitenden Solenoidmagneten erzeugt wird, nach dem Zwiebelschalenprinzip aufgebaut: Die Ausgangspunkte der Teilchenzerfälle werden von einem Silizium-Vertex-Detektor in der Nähe des Strahlrohres vermessen. Auf die Impulse der geladenen Teilchen kann aufgrund ihrer, in der Driftkammer gemessenen, Bahnkrümmungen im Magnetfeld geschlossen werden. Die Identifikation der Teilchen erfolgt mithilfe von Energieverlustmessungen in der Driftkammer, Messungen von Cherenkovstrahlung und Flugzeitbestimmung. Elektromagnetische Schauer werden von einem elektromagnetischen Kalorimeter gemessen und ein System

aus Szintillatoren und Eisenplatten wird zur Identifikation von Myonen und K_L-Teilchen verwendet.

Vollständige Rekonstruktion

Die vollständige Rekonstruktion bezeichnet eine bestimmte Analysetechnik, die ein Hilfsmittel für weitere, physikalische Analysen darstellt. Bei Elektron-Positron-Kollisionen mit einer Schwerpunktsenergie von 10.58 GeV, wie sie am KEK B Beschleuniger gegeben sind, entsteht großteils die $\Upsilon(4S)$-Resonanz, die in ein B-Meson-Paar ohne weitere Zerfallsprodukte zerfällt. Es ist nun das Ziel der vollständigen Rekonstruktion, eines dieser beiden B-Mesonen vollständig zu rekonstruieren. Das heißt, dass alle Spuren und elektromagnetische Cluster, die zu diesem B-Meson gehören ihm zugeordnet werden. Man bezeichnet diesen Teil des Ereignisses als die "Tag-Seite". Dass diese Art der Zuordnung nur bei besonders einfachen Ereignistopologien, wie sie an Elektron-Positron-Beschleunigern vorgefunden werden, möglich ist, verdeutlicht ein Vergleich zwischen den typischen Ereignisstrukturen an Hadron- und Leptonbeschleunigern, der in Abbildung 0.3 gezeigt ist. Der große Nutzen dieser Methode besteht darin, dass wir den Impuls des anderen



(a) Typisches Ereignis am Hadronbeschleuniger, aufgezeichnet vom CDF-Experiment

(b) Typisches Ereignis am Leptonbeschleuniger, aufgezeichnet vom Belle-Experiment

Abbildung 0.3: Vergleich verschiedener Ereignistopologien.

B-Mesons, auf der sogenannten "Signal-Seite" aufgrund der klaren Energieverhältnisse bei den Zerfällen

$$\Upsilon(4\mathrm{S}) \to B^+ \; B^- \quad \text{bzw.}$$
$$\Upsilon(4\mathrm{S}) \to B^0 \; \bar{B}^0$$

genau kennen, ohne eine weitere Rekonstruktion durchführen zu müssen. Desweiteren ist es uns möglich, da wir wissen, dass außer den B-Mesonen keine weiteren Teilchen beim

Zerfall der $\Upsilon(4S)$-Resonanz entstehen, alle Spuren und Cluster, die noch im Detektor gemessen wurden und nicht zur Rekonstruktion auf der "Tag-Seite" verwendet wurden, dem B-Meson auf der Signalseite zuzuordnen. Diese Zuordnung ist zwar im Prinzip bei allen Rekonstruktionen an diesem Experiment möglich, die Besonderheit liegt jedoch in der Zahl der Zerfallskanäle, die innerhalb der vollständigen Rekonstruktion betrachtet werden sollen. Eine vorläufige Liste ist im Anhang B zu finden. Dieses Vorgehen ist natürlich dann besonders sinnvoll, wenn wir an Zerfällen interessiert sind, die sich nur schwer oder gar nicht genau rekonstruieren lassen. Besonders bei Zerfällen, in denen mehrere Neutrinos vorkommen, wie zum Beispiel

$$B^+ \to \tau^+ \, \nu_\tau \to e^+ \, \nu_e \, \nu_\tau \ , \tag{0.5}$$

kann man mithilfe der vollständigen Rekonstruktion Informationen erhalten, die ansonsten nicht zugänglich wären.
Innerhalb der Belle-Kollaboration existieren bereits verschiedene Versionen der vollständigen Rekonstruktion und die Motivation, eine solche Rekonstruktion von Grund auf neu zu entwerfen, liegt in der Hoffnung begründet, mithilfe von neuronalen Netzwerken die Rekonstruktion und Selektion entscheidend verbessern zu können. Schon kleine Verbesserungen bei Teilchen, die in praktisch jedem Kanal in den Listen in Anhang B vorkommen, können die Leistungsfähigkeit der vollständigen Rekonstruktion entscheidend verbessern. Deshalb wird innerhalb des Programms jede Teilchenidentifikation mit neuronalen Netzwerken erneut durchgeführt. Für geladene Teilchen, wie Elektronen, Myonen, Pionen und Kaonen liegen erste Ergebnisse vor, die in den Schaubildern 2.4 bis 2.7 zu sehen sind.
Die Arbeit an der vollständigen Rekonstruktion wurde aufgrund einer aktuellen Veröffentlichung der CDF Kollaboration unterbrochen. Die bis dahin geschaffene Programmstruktur konnte jedoch zur Überprüfung der CDF Analyse verwendet, und dabei gleichzeitig auf Unzulänglichkeiten überprüft werden.

Die Suche nach dem $Y(4140)$

Eine große Vielfalt von gebundenen Zuständen von zwei und drei Quarks ist längst wohlbekannt und ihre Existenz ist experimentell gesichert. Das letzte Kapitel dieser Arbeit widmet sich der Suche nach einem möglichen Kandidaten für einen Vier-Quark-Zustand, der sogenannten $Y(4140)$-Resonanz [T. 09]. Die CDF-Analyse berichtet von einer Struktur im J/ψ ϕ-Massenspektrum im Zerfallskanal

$$B^\pm \to J/\psi \ \phi \ K^\pm \ .$$

J/ψ-Rekonstruktion

Es ist also zunächst vonnöten, das J/ψ zu rekonstruieren, und zwar in den Kanälen

$$J/\psi \to e^+ \, e^- \quad \text{und}$$
$$J/\psi \to \mu^+ \, \mu^-$$

Dazu werden zunächst die Elektron- und Myonkandidaten von jeweils einem neuronalen Netz des NeuroBayes-Pakets klassifiziert (siehe Anhang A.1 und A.2). Danach werden die Myonen und Elektronen jeweils zu den J/ψ-Teilchen kombiniert, wobei wir dabei fordern, dass mindestens eines der beiden Leptonen eine Mindestvoraussetzung bei der NeuroBayes-Klassifizierung erfüllen muss. Dieses Verfahren ist kein Schnitt auf die Leptonkandidaten im üblichen Sinne, da alle Leptonen zur Rekonstruktion verwendet werden, jedoch nicht alle Kombinationen gemacht werden. Dieses Vorgehen resultiert in einer besseren Effizienz bei vergleichbarer Untergrundunterdrückung gegenüber einfachen Schnitten.
Betrachten wir nun noch die J/ψ-Kandidaten, die in Elektronen zerfallen, so tritt hier beim Zerfall ein Strahlungsverlust in Form von emittierten Photonen bei den Elektronen auf. Die zusätzliche Rekonstruktion dieser emittierten Photonen kann das J/ψ Signal verbessern. Wir identifizieren die Photonen mit einem Cluster im elektromagnetischen Kalorimeter innerhalb eines engen Winkelbereichs ($\alpha < 5°$) um die anfängliche Richtung eines Elektrons.

ϕ-Rekonstruktion

Die Rekonstruktion der ϕ-Kandidaten läuft sehr ähnlich ab. Er wird der Kanal

$$\phi \to K^+ \, K^-$$

rekonstruiert. Hier werden die $K^\pm$-Teilchen ebenfalls von einem NeuroBayes-Netz klassifiziert (siehe Anhang A.4) und anschließend zu ϕ Teilchen kombiniert. Wiederum muss mindestens ein $K^\pm$ eine Mindestanforderungen an die NeuroBayes Netzausgabe erfüllen. Auch hier liegt, wie oben erwähnt, kein üblicher Schnitt auf die $K^\pm$ Kandidaten vor.

$B^\pm$-Rekonstruktion

Zur Rekonstruktion der $B^\pm$-Kandidaten werden die Zerfallsprodukte J/ψ, ϕ und $K^\pm$ wiederum durch NeuroBayes-Netze klassifiziert (siehe Anhang A.5, A.6 und A.4). Vor der Rekonstruktion werden die einzigen Schnitte in der Rekonstruktion vorgenommen, und zwar auf den NeuroBayes Ausgabewert der J/ψ-, ϕ- und $K^\pm$-Netzwerke (siehe Tabelle 0.3). Diese Schnitte werden so gewählt, dass sie nur den offensichtlichen Untergrund

Teilchen	NeuroBayes-Schnitt	Untergrundreduktion	Effizienz
J/ψ	-0.980	5.7	99.0%
ϕ	-0.990	8.3	97.6%
$K^\pm$	-0.990	1.7	99.9%
Alle kombiniert		77.6	96.0%

Tabelle 0.3: Schnitte vor der $B^\pm$ Rekonstruktion.

entfernen. Dies dient dazu, die triviale Untergrundkomponente selbst zu entfernen und damit die Qualität der Netzwerkklassifikation auf den restlichen Daten zu steigern. Ein

NeuroBayes-Netzwerk wurde trainiert, um das $B^{\pm} \to J/\psi\ \phi\ K^{\pm}$-Signal vom kombinatorischen Untergrund zu trennen. Das Netzwerktraining beinhaltet unter Anderem Informationen über die Masse der Kandidaten (siehe Anhang A.7), weshalb die finale Selektion der Kandidaten als ΔE-Plot erfolgt, wie beispielhaft in Abbildung 0.4 für einen NeuroBayes Schnitt auf dem $B^{\pm}$ Netzwerk von 0.94 gezeigt. ΔE ist definiert als

$$\Delta E = E_B - E^*_{\text{Beam}} \tag{0.6}$$

wobei E_B die Energie des rekonstruierten B-Mesons ist, und E^*_{Beam} die nominelle Strahlenergie. Für die folgende Analyse wurde der Schnitt von 0.94 weiter verwendet, da er eine

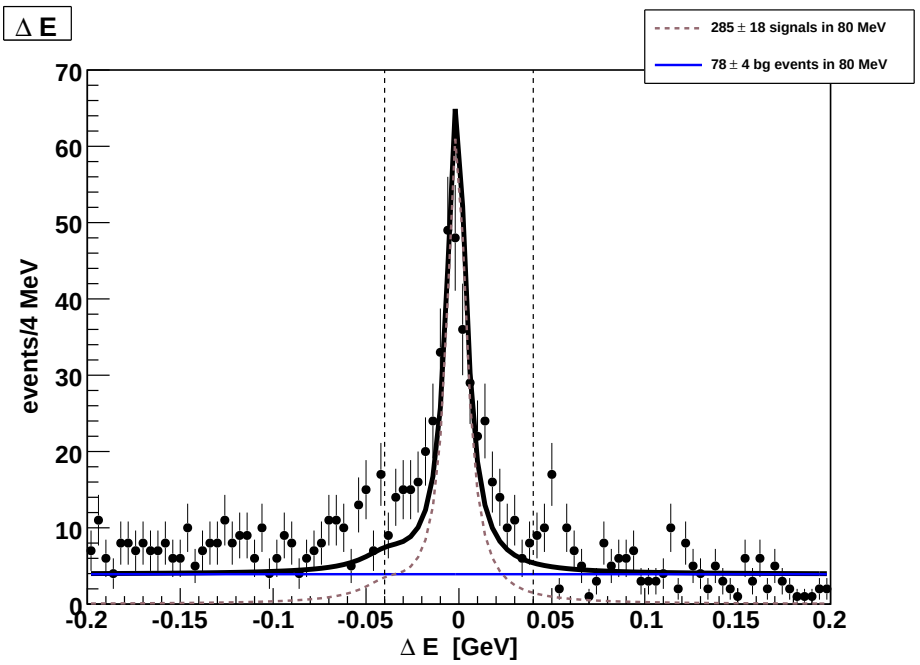


Abbildung 0.4: ΔE Verteilung der Daten.

gute Signifikanz des zu erwartenden Signals liefert. Als finale Selektion wurden die Ereignisse innerhalb eines Bandes von -40 MeV $< \Delta E <$ 40 MeV ausgewählt. Mit diesen Kandidaten wurde schließlich das $J/\psi\ \phi$ Massenspektrum aufgetragen, in dem das Signal vermutet wurde. Es ist in Abbildung 0.5 zu sehen. Die Anpassungsrechnung lieferte ein Ergebnis von 6 ± 6 Signalereignissen. Unter der Annahme der von der CDF-Kollaboration gemessenen Eigenschaften dieser Resonanz, wurden etwa 30 Signalereignisse erwartet. Die Berechnung einer oberen Schranke für das Verzweigungsverhältnis ergab

$$\text{BR}(Y(4140)) < 6.6 \times 10^{-6}\ . \tag{0.7}$$

Die Beobachtung der $Y(4140)$-Resonanz konnte mit den Daten der Belle Kollaboration also nicht reproduziert werden.

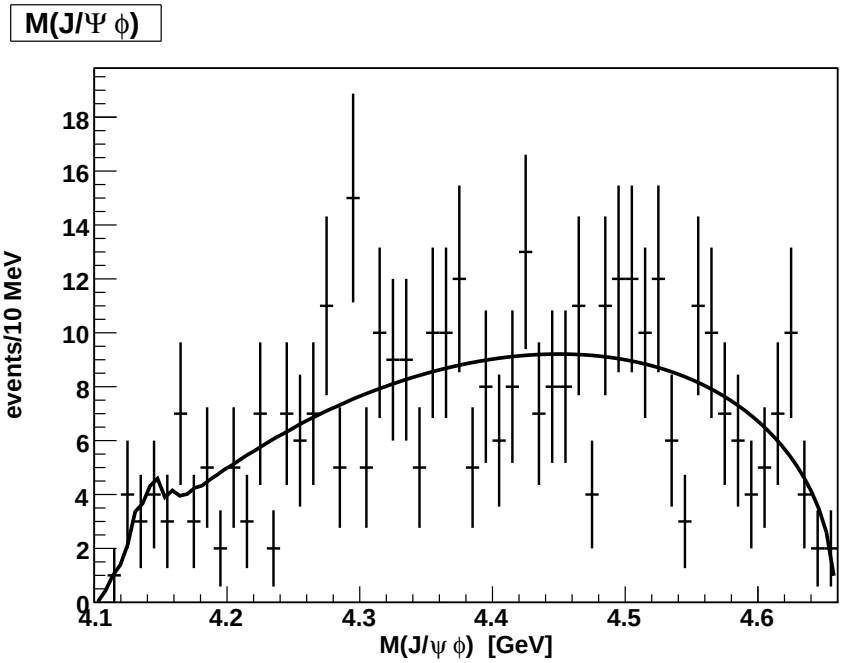


Abbildung 0.5: $J/\psi\ \phi$ Massenspektrum der $B^{\pm} \rightarrow J/\psi\ \phi\ K^{\pm}$ Selektion auf Daten.

Contents

List of Figures

List of Tables

1 Introduction

The theoretical foundation of particles physics is the Standard Model. To continuously test its validity on new phenomena and to measure its free parameters more precisely is the task of the experimental particle physicist. Within this thesis, the foundation for a powerful analysis tool, the full reconstruction, will be developed and an analysis of a candidate for a new kind of resonance, the $Y(4140)$ will be performed.

1.1 A Brief Overview of the Standard Model

Particle physics has to deal, by its very nature, with very small objects and very small distances where the models of classical physics fail. The effects at these small dimensions are described by quantum mechanics, a theory that was created in the early 20th century by amongst others Erwin Schrödinger, Werner Heisenberg and Paul Dirac. The idea of quantisation that was introduced by quantum mechanics can be generalised and leads to quantum field theory, that describes the nature of the particles we know [Per87]. The gauge invariant quantum field theory is able to describe quantum objects and the different forms of interaction between them and is therefore called the *Standard Model of Particle Physics*. The Standard Model is the theoretical foundation for all experimental particle physics and shall therefore be explained below in more detail. The standard model provides the description of all fundamental particles observed so far and the interactions amongst them. Three out of the four known fundamental interactions are described by the standard model via the exchange of *gauge bosons*: The electromagnetic force mediated by the photon (γ), the weak interaction with the $W^{\pm}$ and the Z^0 and the strong interaction with 8 types of gluons. These gauge bosons are particles with spin[1] 1 that mediate the interactions amongst the particles of the standard model. The relative strength of a certain type of interaction is determined by the corresponding coupling constant.

1.1.1 Fermions

Quarks form together with leptons the constituents of matter. Quarks were first proposed in the 1960s to introduce an ordering principle into the large number of hadrons discovered by then. Experimental evidence that supported the quark hypothesis was found in the deep inelastic scattering of electrons on nuclei. Our current knowledge of quarks is that they are without substructure, point-like on a scale of 10^{-17} m and occur in different flavours, which is a way of saying that different quantum numbers

[1] The spin is a property of particles, described by the same theory as regular angular momentum

are assigned to them. These quantum numbers are used in calculations concerning the particle properties and interactions, and therefore help characterising the nature of the particles in question. For historical reasons the quarks are called *down*, *up*, *strange*, *charm*, *bottom* and *top* and can be arranged in three generations, ordered by their mass and their behaviour under the weak interaction. Table 1.1 gives an overview of these three quark generations. Their masses differ largely and bound states with heavy quarks are unstable and decay into bound states containing lighter quarks.
Leptons form the other big group of constituents of matter and occur in pairs of a relatively massive, charged lepton and its corresponding light and neutrally charged neutrino. The massive leptons are called *electron*, *muon* and *tau* and differ only in mass and therefore in lifetime. The electron is stable, as it is the lightest charged lepton and therefore cannot decay into an even lighter one. The muon is heavier than the electron, but not heavy enough to decay into mesons. Muons decay exclusively into electrons. The tau, however is heavy enough to decay into hadrons and there is a large spectrum of decay channels for the tau. *Neutrinos* were first proposed to account for missing energy, momentum and angular momentum in the beta decay of the neutron. To avoid a theory that allowed violation of these properties, a then-undetectable particle was proposed, the neutrino. It was proposed as an electrically neutral and massless particle and was introduced to account for the missing mass and energy in the beta decay. There are three types of neutrinos and each one can be paired up with a massive, charged lepton. Table 1.1 gives an overview of the three lepton pairs.

Family	Name	Charge	Mass
Quarks			
I	up (u)	$+\frac{2}{3}$	$2.55^{+0.75}_{1.05}$ MeV
	down (d)	$-\frac{1}{3}$	$5.04^{+0.96}_{-1.54}$ MeV
II	charm (c)	$+\frac{2}{3}$	$1.27^{+0.07}_{-0.11}$ MeV
	strange (s)	$-\frac{1}{3}$	105^{+25}_{-35} MeV
III	top (t)	$+\frac{2}{3}$	$171.3 \pm 1.3 \pm 1.2$ GeV
	bottom (b)	$-\frac{1}{3}$	$4.20^{+0.17}_{-0.07}$ GeV
Leptons			
I	e	-1	$0.511 \pm 1.3 \cdot 10^{-8}$ MeV
	ν_e	0	< 2 eV
II	μ	-1	$105.7 \pm 0.4 \cdot 10^{-5}$ MeV
	ν_μ	0	< 2 eV
III	τ	-1	1776.8 ± 0.17 MeV
	ν_τ	0	< 2 eV

Table 1.1: Overview of the particles of the standard model [eaPDGon].

1.1.2 Bosons

When describing *interactions*, an important concept is that of conservation. In analogy to classical physics, a law conservation can be linked to an invariance of an interaction. The interaction energy in electrostatics, for example, only depends on the change in the electric potential, not on its absolute value. This invariance is called *gauge invariance* and leads to the conservation of the electric charge. This conservation law also holds in the area of quantum physics. Other conserved properties include charge, momentum and certain quantum numbers, some of which are conserved in all known interactions, others are only conserved by some interactions.

In classical physics, the interaction between two particles is described via a potential or field that is generated by one particle and affects the other one. In quantum theory, however, the concept of the mediation of the three interactions of interest in particle physics is that of an exchange of particles[2] with integer spin, called bosons, associated with the particular type of interaction. Table 1.2 gives an overview over the known bosons. The mentioned three types of interaction are the electromagnetic, the weak and the strong interaction. Particle physics does not deal with the most prominently known interaction, the gravitation, as its effects on processes in particle physics are negligible at the energy scale we are dealing with and there is no consistent quantum field theory available that could include the effects of gravitation into the Standard Model.

Interaction	Name	Charge	Mass
electromagnetic	Photon (γ)	0	0
weak	W^+	$+1$	80.403 ± 0.029 GeV
	W^-	-1	
	Z^0	0	91.1876 ± 0.0021 GeV
strong	Gluon g	0	0

Table 1.2: Overview of the gauge bosons of the standard model [eaPDGon].

1.1.3 The Electromagnetic Interaction

The *electromagnetic* exchange bosons are called photons. They are massless, couple to electrical charge but do not carry an electrical charge themselves. All of the charged particles interact electromagnetically, that is quarks, charged leptons and $W^\pm$ bosons. The coupling constant that characterises the strength of this interaction is also known as the fine-structure constant α and is given by

$$\alpha = \frac{e^2}{4\pi\hbar c} \approx \frac{1}{137} \ . \tag{1.1}$$

[2] A prominent visualisation of this process comprises two people standing on a slippery surface tossing a heavy object, which causes them to repel each other.

The range of this interaction is infinite, as photons are massless and do not interact amongst each other. If we follow the idea of electromagnetic interaction via the exchange of photons for a moment, we can introduce a simple and intuitive way to describe such interactions: the Feynman diagrams. The example shown in figure 1.1 illustrates the process of electron-electron scattering. In the convention I will use in this thesis, the time-flow is in the direction of the positive x-axis and space is measured on the y-axis.

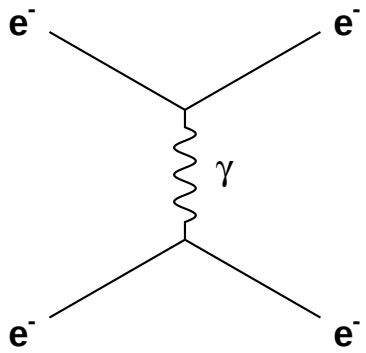


Figure 1.1: Electron-electron scattering.

1.1.4 The Strong Interaction

Quarks interact by the *strong* interaction that is described by quantum chromodynamics. The exchange bosons of this interaction are called gluons. Gluons are massless and do not carry electrical charge. They couple to the strong charge that we call colour and also carry colour themselves, so they also interact with each other. In the Standard Model we assume three colours and their respective anti-colours that we call (anti-)red (r,$\bar{\text{r}}$), (anti-)blue (b,$\bar{\text{b}}$) and (anti-)green (g,$\bar{\text{g}}$). The gluons form a colour octet whose composition is somewhat dependent on convention. An example [Per87] is:

$$\text{r}\bar{\text{b}} \qquad \text{r}\bar{\text{g}} \qquad \text{b}\bar{\text{g}} \qquad \text{b}\bar{\text{r}} \qquad \text{g}\bar{\text{r}} \qquad \text{g}\bar{\text{b}} \qquad \frac{\text{r}\bar{\text{r}} - \text{b}\bar{\text{b}}}{\sqrt{2}} \qquad \frac{\text{r}\bar{\text{r}} + \text{b}\bar{\text{b}} - 2\text{g}\bar{\text{g}}}{\sqrt{6}} \; .$$

Evidence for the internal quantum number "colour" comes from the existence of the Δ^{++} particle, that is measured as a bound state of three up-quarks in the spatially symmetric ground state with angular momentum $l = 0$ with the spins coupled to $J = \frac{3}{2}$, which can figuratively be described as "all three spins parallel". So the wave function of this particle is symmetric in flavour, space and spin, which seems to be a violation of the Pauli principle, which states that multiple fermions cannot exist in the same quantum state. Hence, an additional quantum number was introduced resolving these difficulties: the colour. Experimental evidence shows that quarks do not occur as free particles, but only confined, that is in bound states of either two quarks (called *mesons*) or three quarks (called *baryons*) that we call colourless[3]. In these bound states either two quarks

[3] An exception might be the top quark, as it does not live long enough to form a bound state

carry colour and anti-colour respectively, which results in a colourless state, or three quarks carry all three different colours, which again results in a colourless state. We can see the analogy that superimposed red, blue and green light results in white light.

$$\text{red} + \text{anti-red} = \text{colourless} \qquad \text{(Meson)}$$
$$\text{red} + \text{green} + \text{blue} = \text{colourless} \qquad \text{(Baryon)}$$

Quantum chromodynamics also allows other possibilities for colourless states, such as four-quark or five-quark bound states, which we summarise under the term "exotic hadrons":

$$\text{red} + \text{anti-red} + \text{green} + \text{anti-green} = \text{colourless} \qquad \text{(Tetraquark)}$$
$$\text{red} + \text{green} + \text{blue} + \text{red} + \text{anti-red} = \text{colourless} \qquad \text{(Pentaquark)}$$

There have been numerous claimed discoveries of tetra- and pentaquarks, but most of these observed signals could be explained satisfactorily without the need for exotic hadrons or turned out to be non understood background.
The form of the strong potential offers an explanation for this *quark confinement*: The strong potential that can be assigned to a pair of quarks is of the form

$$V_S = -\frac{4}{3}\frac{\alpha_S}{r} + kr \ . \tag{1.2}$$

The first term ($\sim 1/r$) is clearly dominant at small ranges and proportional to the Coulomb potential in electrodynamics. At large distances, however, the second term ($\sim r$) dominates clearly and gives an explanation for the above mentioned fact that quarks only occur in bound states of two or three quarks. Let us take two quarks as an example and try to increase the distance between them in order to create two free quarks. According to the potential given in equation 1.2 the amount of energy necessary to do so grows in such a rapid manner, that it is energetically favourable to create a new quark-anti-quark pair. Each of the newly created quarks can form a new meson together with one of the original quarks. This process is called hadronisation. The strong force is, as its name suggests, the strongest interaction in nature with a coupling constant about 100 times larger than the electromagnetic coupling constant

$$\frac{\alpha_S}{\alpha} \approx 100 \qquad \Rightarrow \qquad \alpha_S \approx 1 \ . \tag{1.3}$$

The coupling of the colour of quarks and gluons is shown schematically in figure 1.2.

1.1.5 The Weak Interaction

The only interaction in which all quarks and leptons, including the neutrinos, participate in is the *weak* interaction. As its name suggests, its effects are so small compared to the electromagnetic and the strong interaction that it only plays an important role, when those interactions are suppressed by conservation laws. The gauge bosons of the weak interaction are the electrically neutral Z^0 boson and the two charged W^+ and W^-

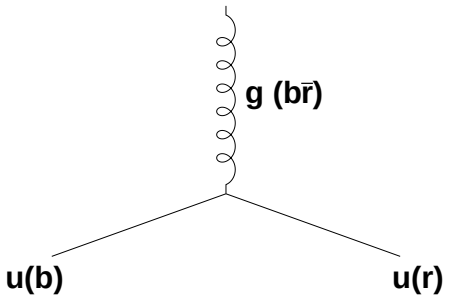


Figure 1.2: Colour coupling.

bosons. The phenomena that we now explain with the help of the Z^0 and $W^{\pm}$ bosons were historically called neutral and charged currents respectively. The Z^0 and the $W^{\pm}$ carry a high mass and decay after a short lifetime which results in the small range of the weak interaction. The weak coupling constant can be obtained by comparing decays initiated by the weak and the strong interaction respectively:

$$\alpha_W \sim 10^{-6} \ . \tag{1.4}$$

A prominent and historically also the first example for a weak process is the β-decay of the neutron, in which the neutron $n(ddu)$ decays into a proton $p(uud)$, an electron e^- and an anti-electron neutrino $\bar{\nu_e}$:

$$n \rightarrow p^+ e^- \bar{\nu_e} \ . \tag{1.5}$$

The Feynman graph to this decay is shown in figure 1.3 As it can be seen in the β decay,

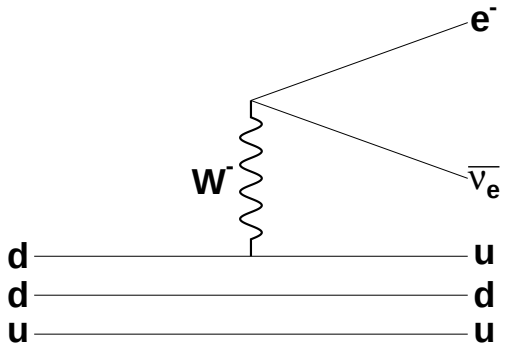


Figure 1.3: β decay.

the weak interaction is capable of changing the quark flavour and it is the only interaction with that capability. To be more precise, the charged current $W^{\pm}$ interactions change the quark flavour. So the decay of heavy quarks to light quarks must be a weak process

and the involved $W^{\pm}$ mesons couple to the up- and down-type quarks with relative strengths given by the Kobayashi-Maskawa Matrix [eaPDGon]:

$$\begin{pmatrix} V_{u\to d} & V_{u\to s} & V_{u\to b} \\ V_{c\to d} & V_{c\to s} & V_{c\to b} \\ V_{t\to d} & V_{t\to s} & V_{t\to b} \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.226 & 0.004 \\ 0.226 & 0.973 & 0.042 \\ 0.009 & 0.041 & 0.999 \end{pmatrix} \ . \tag{1.6}$$

1.2 Experiments

The matter that surrounds us in everyday life consists only of leptons and three-quarks states of the first family. First evidence that there could exist other particles than electrons, protons and neutrons was found in the analysis of cosmic particles. Neutrinos, muons, pions and kaons were discovered in these experiments. These newly found particle did not come from another place in the universe, where there are other forms of matter, but they were *produced*. This production of particles can be understood with the help of Albert Einstein's infamous formula

$$E = mc^2 \ , \tag{1.7}$$

which states that energy and mass can be transformed into one another. The creation of new particles out of energy and the decay of heavy particles into lighter ones via emission of energy are crucial processes for particle physics.
The energy densities required to produce the heavier quarks, the charm-, bottom- and top-quarks can, under controlled conditions, only be achieved in modern particle accelerators. Although the several machines in existence differ largely, they share the common principle of accelerating stable, charged particles using electric fields and colliding the accelerated particles either with a fixed target or with other accelerated particles moving in the opposite direction. An obvious difference is the achievable centre of mass energy $\sqrt{s}$ [Hin97]. The historically seen first experiments were conducted with fixed targets, an experimental setup where the accelerated particles with energy E_1 and momentum p_1 are directed onto a fixed target with energy m_2 and no momentum, usually a sample of matter whose microscopic properties were to be examined. The same setup can also be used to reach high energies and create new particles, but the achievable energy in the centre of mass frame $\sqrt{s}$ is

$$s = 2E_1 m_2 + m_1^2 + m_2^2 \ , \tag{1.8}$$

$$\sqrt{s} \sim \sqrt{E_1} \ . \tag{1.9}$$

For a collider with two relativistic particles with energies E_1, E_2 that are colliding head-on with momenta p_1, p_2, the centre of mass energy $\sqrt{s}$ can be calculated as:

$$s = 2(E_1 E_2 + p_1 p_2) + 2M^2 \approx 4E_1 E_2 \ , \tag{1.10}$$

$$\sqrt{s} \sim E \ , \tag{1.11}$$

where $E^2 = E_1 \cdot E_2$. We can see that the available energy for the creation of new particles grows with $\sqrt{E}$ for fixed-target experiments, but with E for collider experiments, so in

order to reach high energies, collider experiments are the way to go.
Modern colliders can be categorised by the types of particles that are accelerated:

Electron-positron colliders have the advantage to utilise particles that have no substructure, so the available centre of mass energy in a collision can be chosen precisely. The largest drawback of electron-positron colliders is the huge amount of synchrotron radiation they produce. This type of radiation is emitted by any charged particle under acceleration and has the consequence of energy loss ΔE of the electrons and positrons in a collider. The energy loss for circular acceleration can be expressed as

$$\Delta E = \frac{1}{3}\left(\frac{e^2\beta^3\gamma^4}{\rho}\right) , \tag{1.12}$$

where ρ is the bending radius, β the velocity of the charged particle and $\gamma = (1-\beta^2)^{-1/2}$. With $\beta \sim |\vec{p}|/m$, it becomes clear that the lighter the accelerated particle is, the larger its energy loss by synchrotron radiation is:

$$\delta E \sim \frac{1}{m^4} \tag{1.13}$$

This synchrotron radiation was and still is the major energy loss for electron-positron colliders like LEP, where a single particle with a nominal energy of ~ 100 GeV lost an energy of ~ 2.5 GeV per turn.

Hadron colliders usually accelerate protons or protons and anti-protons and bring them to collision. As the energy loss due to synchrotron radiation is strongly mass dependent, according to equation 1.13 we can easily see that this energy loss is a factor of $(m_p/m_e)^4 \approx 10^{13}$ smaller for protons than for electrons. Therefore much higher energies can be reached with hadron colliders. The disadvantage of these colliders is the substructure of the accelerated particles. Usually, only one constituent of each hadron plays an active role in the collision, and this constituent carries only a small fraction of the entire energy of the hadron. As this fraction varies strongly, a precise choice of energy is impossible. Nevertheless, higher energies can be reached for the interacting constituents with hadron colliders.

As a conclusion, one could say that hadron colliders are more suited to achieve high energies and electron-positron colliders are more suited to perform high precision measurements. High energy accelerators, like the LHC and the Tevatron, yield enough energy to push the boundaries of achievable energies and thereby the potential to discover new particles further, whereas lepton colliders such as LEP, PEP-II or KEK B can achieve much higher precisions and are therefore more suited to perform high-precision measurements that also allow the discovery of new physics and indirectly also new fundamental particles.

1.3 The KEK B Accelerator

The KEK B accelerator is an asymmetric electron-positron collider [Rep95] at Tsukaba, Japan, in the northwest of Tokyo. Its construction began in 1994 in the underground tunnels that were formerly occupied by the TRISTAN collider. Figure 1.4 shows a schematic drawing of the KEK B accelerator. KEK B consists of two rings, constructed

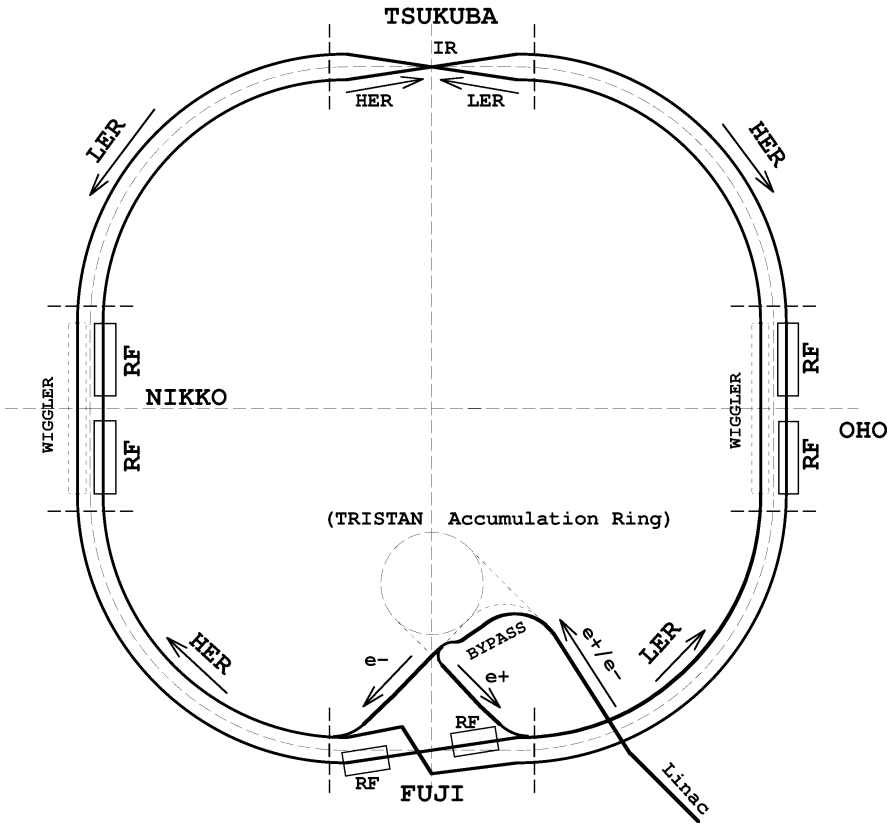


Figure 1.4: The KEK B accelerator.

side by side: A low energy ring (LER) that contains a positron current of approximately 1600 mA [Rep09] at an energy of $E_+ = 3.5$ GeV and a high energy ring (HER) that contains an electron current of approximately 1200 mA at an energy of $E_- = 8.0$ GeV. The asymmetry in the energy of these two rings is derived from physics considerations, as one of the main goals of the experiments is to measure CP violations in the decay of B mesons, for which B mesons with finite momenta have to be produced that travel measurable distances before they decay and thus enable the measurement of their decay pattern.

The two beams with energies $E_+ = 3.5$ GeV and $E_- = 8.0$ GeV yield a centre of mass

energy $\sqrt{s}$ of 10.58 GeV, which is approximately the energy of the $\Upsilon(4S)$ resonance:

$$\sqrt{s} = \sqrt{4E_+E_-} \approx 10.58 \text{ GeV} = E_{\Upsilon(4S)} \ . \tag{1.14}$$

The asymmetry between the two rings can be expressed by the resulting Lorentz-boost parameter $\beta\gamma$ of the $\Upsilon(4S)$ resonance:

$$\beta\gamma = \frac{E_- - E_+}{\sqrt{s}} = 0.42 \ . \tag{1.15}$$

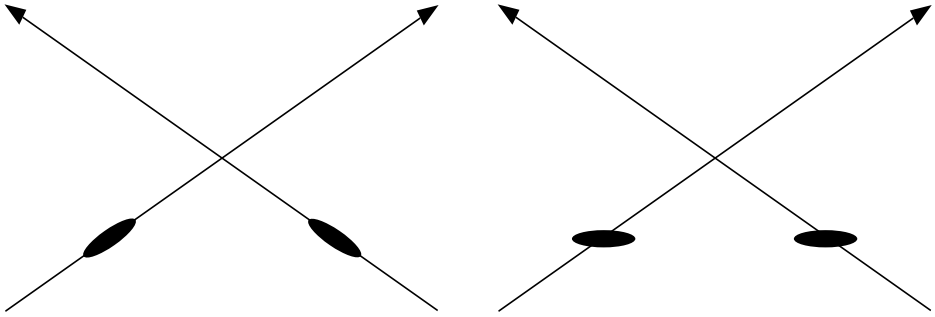
Although we speak of a "particle beam", the shape of the particles in a collider is not a constant flow, but rather thousands of particle bunches, a few centimetres in length with equal spacing of approximately 2m between them. This topology is due to the mechanism of acceleration by high frequency electromagnetic fields that self-actingly produces this bunch structure.
The production rate for new particles in collisions is determined not only by the achieved energy, which is only an indicator if a certain type of particle can be produced at all with the available energy, but by the *luminosity* of the collider. The luminosity of a collider depends only on the properties of the beam and is given by

$$\mathcal{L} = \frac{N_1 N_2 f}{4\pi\sigma_x\sigma_y} \ , \tag{1.16}$$

where $N_{1,2}$ is the number of bunches in the HER and LER ring respectively, σ_x and σ_y are the spatial dimensions of the bunches and f is the collision rate of these bunches. The KEK B accelerator has 5120 buckets, which are places where bunches could be. These buckets are filled with trains of 1582 bunches, followed by a gap of 270 buckets, which is needed for a safe beam abort procedure. The bunch crossing frequency is at $f \approx 508.9$ MHz and bunches are crossed at an angle of 22 mrad in the horizontal plane. In order to keep the effective crossing area $4\pi\sigma_x\sigma_y$ low and thus increasing the luminosity, crab cavities are used, that rotate the bunches in such a manner that, in spite of the finite crossing angle, they collide head on, as shown in figure 1.5. The KEK B has only got one interaction region, which lies in the Tsukuba area, where the Belle detector is located (see chapter 1.4). The KEK B accelerator was designed for a luminosity of $1 \cdot 10^{-34}$ $\text{cm}^{-2}\text{s}^{-1}$, but as of June 2009 has reached a record luminosity of $2.1083 \cdot 10^{-34}$ $\text{cm}^{-2}\text{s}^{-1}$ and is currently the machine with the highest luminosity in the world.

1.4 The Belle Experiment

The *Belle detector* is a particle detector designed to perform high precision measurements of asymmetric electron-positron collisions at the $\Upsilon(4S)$ resonance [ea95]. Figure 1.6 shows a schematic side view of the detector. Belle is constructed around a superconducting solenoid which provides a magnetic field of 1.5T. The vertices of particle collisions are measured using the Silicon Vertex Detector (SVD) near the beam pipe, the momenta of charged particles are measured with the Central Drift Chamber (CDC) via



(a) Beam crossing at finite angle (b) Beam crossing with crab cavities

Figure 1.5: Beam crossing.

the curvature of their reconstructed tracks. Particle identification is provided by dE/dx measurements in the CDC, an Aerogel Cherenkov Counter and a Time Of Flight (TOF) system. Electromagnetic showers are measured with an Electromagnetic Calorimeter (ECL) and Muons and K_L Mesons are detected by resistive plate counters. These above mentioned system shall be described below in more detail. The coordinates used for the description of the Belle detector are shown in figure 1.7: The positive z-axis points in the direction of the high energy beam, the x-axis is oriented perpendicular to the z-axis and lies in the horizontal plane, the y-axis is perpendicular to the x- and z-axis. φ is measured between the x- and y-axis and θ is measured between the z- and y-axis.

Beam pipe Although it is strictly speaking not a component of the detector, the beam pipe nevertheless has a certain influence of the way Belle works, as multiple Coulomb scattering can occur in the beam pipe wall. An additional issue is the intense heating of the beam pipe in the interaction region ($\sim$ 100W), while the first layers of the actual detector should be placed as close to the beam pipe as possible without being damaged by the heat. So the configuration of a thin $d = 0.5$ mm beryllium inner and outer wall with a 2.5 mm helium gas layer for cooling between the two walls was chosen. The beam pipe is of cylindrical shape and has an inner diameter of of $d = 40$ mm.

SVD2 The Silicon Vertex Detector 2 is used to measure the distances of the decay vertices of B mesons, a measurement that is essential for the study of time dependent CP asymmetries, one of the most prominent goals of Belle. The SVD2 also provides decay vertex identification and measurement not only for B mesons, but also for D mesons and τ leptons. Furthermore it contributes to the tracking of particles. The SVD2 is the innermost component of the Belle detector [Gro01]. It is the successor of the original SVD, that was replaced due to massive radiation damage. The SVD2 consists of five layers of double sided silicon strip detectors with the following radii:

$$r_1 = 15.0 \text{ mm}, \quad r_2 = 21.1 \text{ mm}, \quad r_3 = 44.0 \text{ mm}, \quad r_4 = 70.0 \text{ mm}, \quad r_5 = 90.0 \text{ mm} .$$

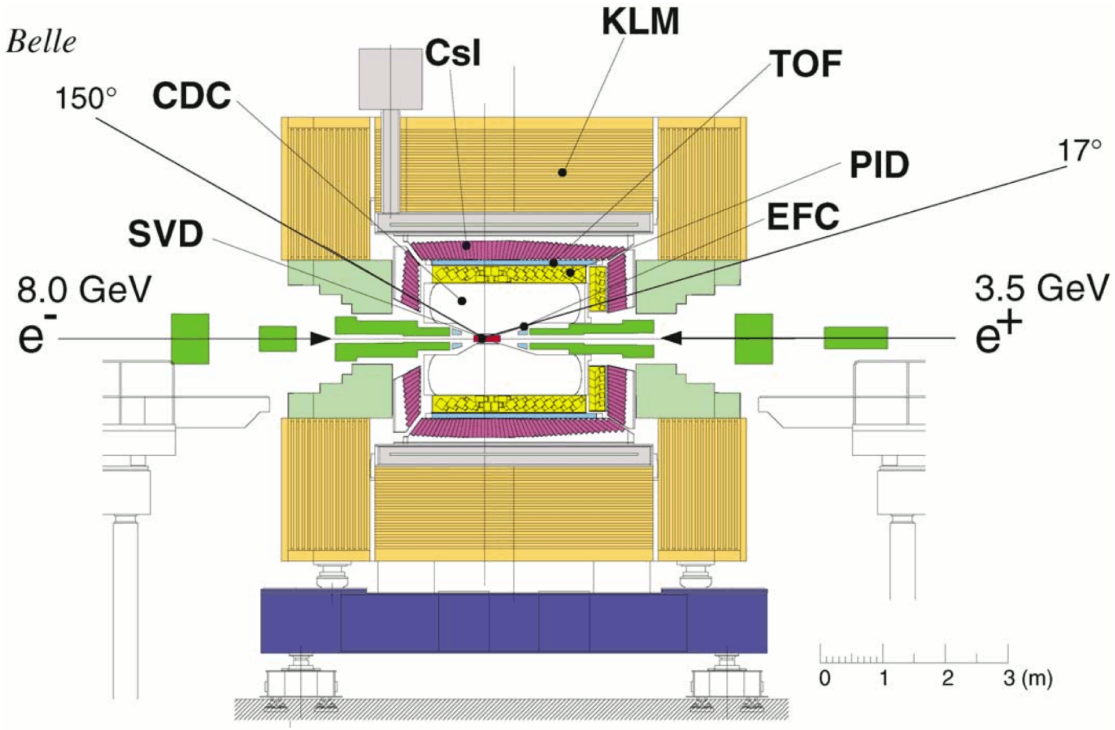


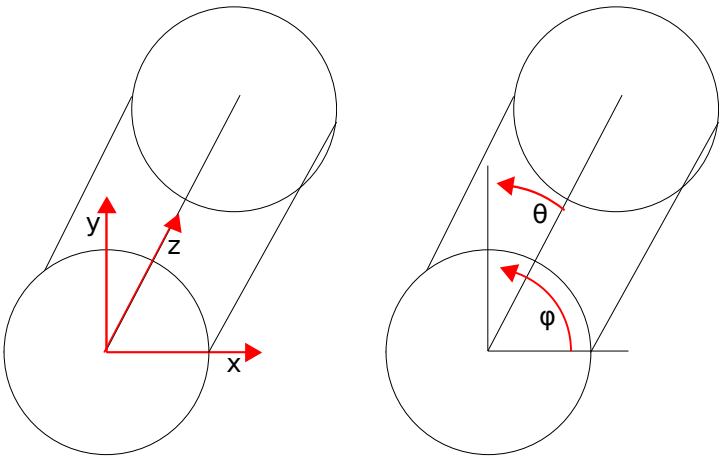
Figure 1.6: Side view of the Belle detector.

It provides an angular coverage of $17° < \theta < 150°$. The precision of the z-difference measurements is

$$\sigma_{\Delta z} \sim 80\ \mu\text{m}\ .$$

CDC The Central Drift Chamber is designed to fulfil the important role of reconstructing charged particle tracks, which allow by measurement of their curvature conclusions about the momenta of the associated particles. Additionally dE/dx information can be obtained from the drift chamber, which contributes to the identification of the observed particle.
The CDC consist of 50 layers of either axial or stereo wires and they are configured cylindrically around the beam axis, asymmetric in the z-direction to provide an angle coverage of $17° < \theta < 150°$. This configuration results in a total of 8400 nearly square drift cells with a maximum drift distance of $8 - 10$ mm. The whole drift chamber is filled with a mixture of 50% helium and 50% ethane gas to minimise multiple Coulomb scattering and obtain a good dE/dx resolution at the same time. The CDC delivers the following resolutions:

$$\sigma_{r\phi} = 130\ \mu\text{m} \qquad \sigma_z = 200 - 1400\ \mu\text{m} \qquad \frac{\sigma_{p_t}}{p_t} = \frac{0.3\%}{\sqrt{p_t^2 + 1}} \qquad \sigma_{\text{dE/dx}} = 6\%$$



(a) x,y and z direction (b) φ and θ direction

Figure 1.7: Geometry explanation.

ACC The Aerogel Cherekov Counters are installed just outside of the CDC and provide additional information for the identification of particles. A main goal is the notoriously difficult task of separating $K^{\pm}$ from $\pi^{\pm}$.
The ACC consist of 960 counter modules in the ϕ direction and 228 modules for the forward direction. The counters are read out by fine mesh-type photomultiplier tubes that deliver a certain absolute number of photoelectrons $N_{\text{p.e.}}$ that is used to determine the particle type. A good K/π separation can be achieved for momenta of 1.2 GeV/$c < p <$ 3.5 GeV/c.

TOF The Time-Of-Flight detector component is used to measure the time the particles needed to travel from the interaction point to the TOF scintillators. It thereby gives additional information for the identification of particles. The TOF is located outside the ACC, at a distance of approximately 1.2 m from the interaction point. It utilises scintillators with a time resolution of $\sim$ 100 ps. The 64 TOF modules cover a polar angle of $34° < \theta < 120°$ and are read out by fine-mesh-dynode photomultiplier tubes. A good K/π separation can be achieved for momenta $p <$ 1.2 GeV/c.

ECL The Electromagnetic Calorimeter of the Belle detector is primarily used to measure energy and position of photons. Those photons can either stand at the end of a decay chain and have relatively low energies ($<$ 500 MeV), but also some B-decay modes produce photons as direct decay products, such as $B \to K^*\gamma$ which requires good performance of the ECL up to 4 GeV. For the identification of π^0 particles, a good angular resolution is needed.

The ECL is composed of CsI(Tl) crystals, with a radiation length of 1.85 cm, that are read out by a silicon photodiode. It stretches from a radius of 125 cm to 162 cm and from -102 cm to 196 cm in the z-axis. It consists of three large segments:

- the forward end-cap with an angular coverage of $12.4^\circ < \theta < 31.4^\circ$ consisting of 1152 crystals,
- the barrel with an angular coverage of $32.2^\circ < \theta < 128.7^\circ$ consisting of 6624 crystals and
- the backward end-cap with an angular coverage of $130.7^\circ < \theta < 155.1^\circ$ consisting of 960 crystals.

The entire system contains 8736 CsI(Tl) counters and weighs 43 tons. Its resolutions in energy and position are

$$\frac{\sigma_E}{E} = \frac{1.3\%}{\sqrt{E}} \qquad\qquad \sigma_{\text{pos}} = \frac{0.5 \text{ cm}}{\sqrt{E}} \ .$$

EFC The extreme forward calorimeter improves the angular coverage of the ECL, covering $6.4^\circ < \theta < 11.5^\circ$ in the forward direction and $163.3^\circ < \theta < 171.2^\circ$ in the backward direction. As it lies very close to the interaction region, it has to be radiation-hard. Therefore a Bismuth-Germanate ($Bi_4Ge_3O_{12}$) crystal calorimeter was chosen that provides an energy resolution of $(0.3 - 1)\%/\sqrt{E(\text{ GeV})}$.

KLM The K_L and Muon detection system is composed of alternating layers of 15 detectors for charged particles and 14 iron plates of 4.7 cm thickness providing a total of 3.9 interaction lengths for particles travelling perpendicular to the normal direction. The incoming K_L particles interact within the iron plates and produce showers of ionising particles, whose location indicate the direction of the K_L. The charged particle detectors allow separation between muons and charged hadrons ($K^\pm$ and $\pi^\pm$) with muons travelling much further with smaller deflections.

Magnet The magnetic field required by Belle to determine the momenta of charged particles via their curvature in the CDC is produced by a superconducting NbTi/Cu solenoid that provides a magnetic field of 1.5 T. The coil is of cylindrical shape with an effective radius of 1.8 m and a length of 3.92 m. It typically takes half an hour to charge the coil before it reaches its nominal current of 4400 A, which corresponds to an energy of 35 MJ.

Planning for the Belle experiment started in 1994, the design report was written in 1995 and after some design changes in 1997, Belle was commissioned in May 1999.

1.5 Data Analysis

1.5.1 Probability

When we look at classical physical processes, like a pendulum or the orbits of planets, the outcome of a measurement is exactly predictable. There is no fundamental uncertainty, a defined cause has a defined effect. This is the concept which we call determinism. The other extreme would be a process whose outcome is completely non-predictable, like the exact time of a radioactive decay of a single atom or the outcome of a lottery, which we call coincidence.
Most real systems are, however, a mixture of determinism and pure coincidence [BL98]. The classical expectation, that a certain experiment should give, under identical circumstances, identical results is an idealisation even in classical physics, as the circumstances of two experiments can never be exactly identical. We find an even worse situation in the field of quantum mechanics, where an experiment, even if the circumstances could be identical, would never yield exactly predictable results. This behaviour is not due to our current lack of experimental finesse, but an essential property of quantum mechanics and thereby of particles physics, ultimately due to Heisenberg's uncertainty principle. It is therefore impossible for any physicist, and for any particle physicist in particular, to measure any quantity exactly and we are limited to make statements about the *probability* of the outcome of a measurement. We define the probability $P(A)$ for the outcome A of an experiment with n independent measurements and k positive outcomes as:

$$P(A) = \lim_{n\to\infty} \frac{k}{n} \, . \tag{1.17}$$

A typical example would be the probability of an observed particle to have a mass within certain limits or the probability of a particle to be a π^- or an e^-. The definition in equation 1.17 yields one practical problem, which is that the infinite number of repetitions of the experiment can never be achieved. It can however be shown that the error on the probability can be reduced as far as desired by increasing the number of repetitions.
An important property of probabilities is the ability to make combinations. Let the probability of A be $P(A)$ and the probability of B be $P(B)$, then the probability that A and B occur is

$$P(A \text{ and } B) = P(A) \cdot P(B|A) \tag{1.18}$$

where $P(B|A)$ is the conditional probability that B occurs, given that A has occurred. As A and B are not specified any further we can also write in general

$$P(A \text{ and } B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \tag{1.19}$$

which leads to Bayes' Theorem:

$$P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)} \, . \tag{1.20}$$

This theorem will be needed for the statistical analysis performed in this thesis, which will be explained in section 1.6.

1.5.2 Distributions

As mentioned in section 1.5.1 most measurements in nature, and in particular in particle physics, do not yield one fixed and predictable value, but rather a distribution of values. Figure 1.8 shows a measurement of a particle mass, which is a typical distribution in particle physics. We can intuitively see a peaking structure and a downward sloped

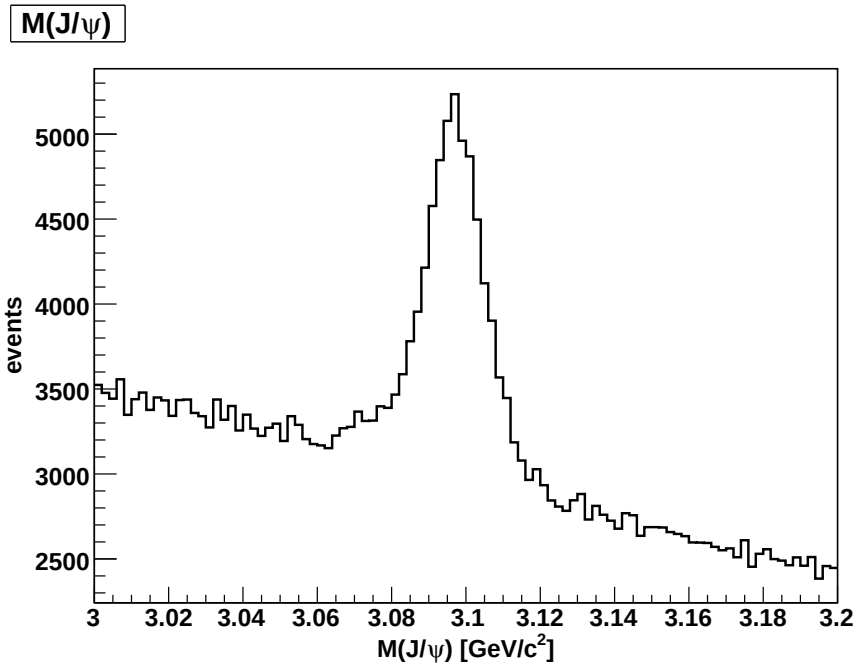


Figure 1.8: A typical mass distribution.

flat component. The peak can be identified as mass measurements of the particle we are interested in (J/ψ in this example), which we call signal and the downward sloping component can be identified as mass measurements of other particles whose masses were close or identical to the mass of the signal component, but were not signals in reality. We call this component the background.
When we measure a continuous quantity like mass, it is in principle impossible to measure one exactly predicted value (say $m = 3.096$ GeV/c^2). Therefore we introduce the *probability density function (pdf)* $f(x)$, a property that characterises the distribution of a random variable. It denotes the probability that a measurement of the quantity x lies within the limits a and b by

$$P(a \leq x \leq b) = \int_a^b f(x)\mathrm{dx} \ . \tag{1.21}$$

The application of this formula can be shown in the exemplary plot in figure 1.8: The probability that the outcome of a mass measurement will lie within the peaking structure is higher than within an equally wide range outside the peak. Note that the probability

density function $f(x)$ is not a probability, but rather the quantity $f(x)\Delta x$ will become the probability for x to lie within x and Δx as $\Delta x \to 0$. This makes clear, that x cannot take an exact and predefined value.

1.6 Classification

During a typical high energy physics analysis, one is often presented with a collection of reconstructed particles, called *candidates*, some of which are actually the particle one is looking for, which we call the *signal component*, and some of which are not the particles one is interested in, which we call *background*. In order to obtain results from an analysis, it is essential to separate signal from background. For all selection methods, a compromise has to be found between good *efficiency* and good *purity*, which can be defined as:

$$\text{efficiency} = \frac{\text{N(selected signals)}}{\text{N(signals in the sample)}}, \qquad \text{purity} = \frac{\text{N(selected signals)}}{\text{N(selected candidates)}} \; . \tag{1.22}$$

One way to influence these quantities is to change the values of certain cuts. For a more detailed explanation of cut based analyses, see chapter 1.6.1. Increasing the efficiency by applying softer cuts usually results in more background passing those cuts, lowering the purity, whereas choosing harder cuts, on the other hand, will improve the purity by cutting away more background but will also lower the efficiency by cutting away a relatively large signal component. Figure 1.9 shows an example of purity plotted

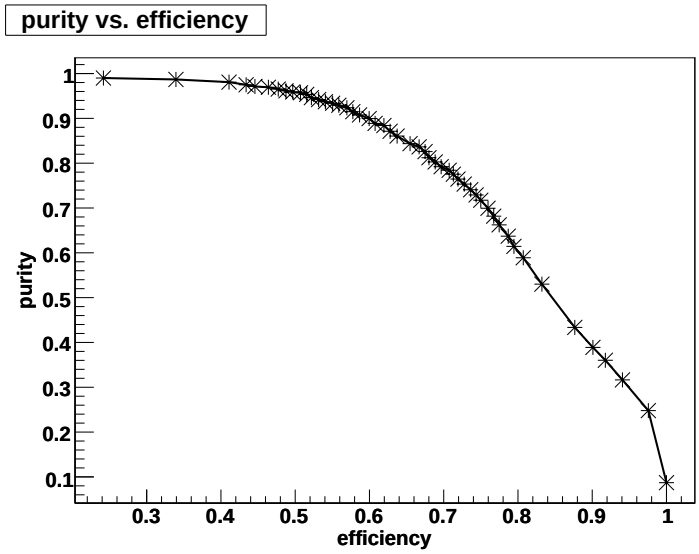


Figure 1.9: An exemplary purity-efficiency-plot.

over efficiency. The curve comprises all possible settings of purity and efficiency for a given classification procedure. This plot can be used as a quality characteristic for the

used procedure with better classification reaching further into the top right corner of the graph. Having this in mind, it is crucial for any analysis to choose a good way of classification.

1.6.1 Cut based analysis

The easiest way to separate signal from background is to apply sequential cuts on the properties of a candidate, for example on the mass, momentum or impact parameter. Using this procedure it is very difficult to take correlations of the variables into account, which could result in low efficiencies or purities. Figure 1.10 gives a visualisation of this method. Let us assume, that the black region is the desired signal, the x- and y-axis are

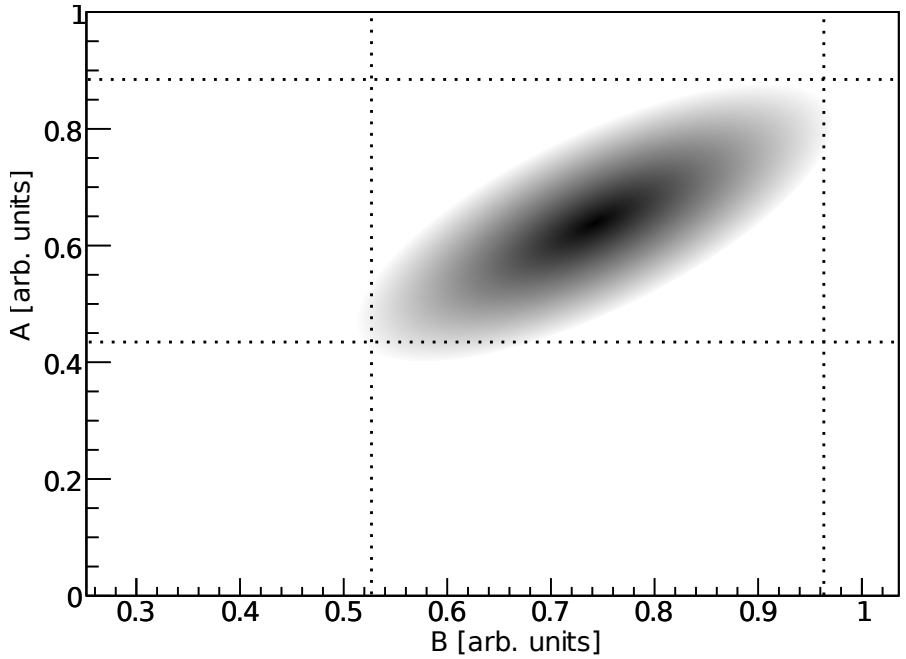


Figure 1.10: Visualisation of sequential cuts (dotted lines), black: signal.

properties of the candidates and let us further assume that the background is equally distributed ("flat"). Cuts on the properties represented by the x- and y-axis are visualised by the black dotted lines. We can see, that it is unavoidable to either cut away a lot of the signal to achieve a good purity or to leave a lot of the background in to achieve a good efficiency.

1.6.2 Multivariate Techniques

A common technique to achieve more sophisticated cuts is to combine all significant variables available into on scalar variable, for example a likelihood ratio, and to perform a cut on this new variable. These multivariate techniques are in principle capable of

taking correlations of the variables into account. The application of these techniques can, however, be rather involved. Simplified models can deliver quite good results when correlations between the different variables are small.
Another example of a multivariate technique is a neural network. The neural network that was used prominently in this thesis is part of the NeuroBayes package, that shall be described below in more detail.

1.6.3 NeuroBayes

The idea of the NeuroBayes package is to pass all of the relevant variables through a preprocessing algorithm to a neural network [Fei04]. This network is capable of operating in two different modes, which is either classification- or density-mode.

- The classification mode is the most frequently used mode in particle physics and also the only mode used in this thesis. For a classification task, which is usually the decision, if a candidate is signal or background, the network's task is to map the input variables to a single output variable, while taking into account the correlations of the input variables. The produced output is a measure of probability for the candidate to be a signal and can be transformed easily into a real probability in the Bayesian sense (see equation 1.20). A single output node is used for this measure of probability.

- In density mode, there is not one, but many output nodes, which enable the network to estimate a probability density function, from which one can obtain estimates for absolute values rather than just a probability.

As the NeuroBayes package, and its classification mode in particular, is used extensively in the analysis described in this thesis, it will be explained in more detail:

Theory

Let us assume that we have a random variable t distributed according to its probability density function (pdf) $f(t)$ (see section 1.5.2). In particle physics, our knowledge of the pdf will usually come from Monte Carlo simulations. If the signal can be seen clearly in a distribution from real data, the pdf can be also estimated from a large number of measurements. The pdf of the measurement shown in figure 1.8 can be estimated easily. Given the knowledge of this shape, it is easy to say, what outcome a measurement will have *on average*. NeuroBayes, however, will aim to produce a better estimate for *each single event* taking into account a set of measured values $\vec{x_i}$. This means that NeuroBayes aims to find the conditional probability density $f(t|\vec{x_i})$. If there is no information in the measured values $\vec{x_i}$ for the current classification problem, then our best estimate is the pdf: $f(t|\vec{x_i}) = f(t)$. If there is, however, information in the input vector $\vec{x_i}$, one should be able to obtain a better estimator for a given event with the conditional probability density $f(t|\vec{x_i})$ than with the pdf $f(t)$ alone.

Preprocessing

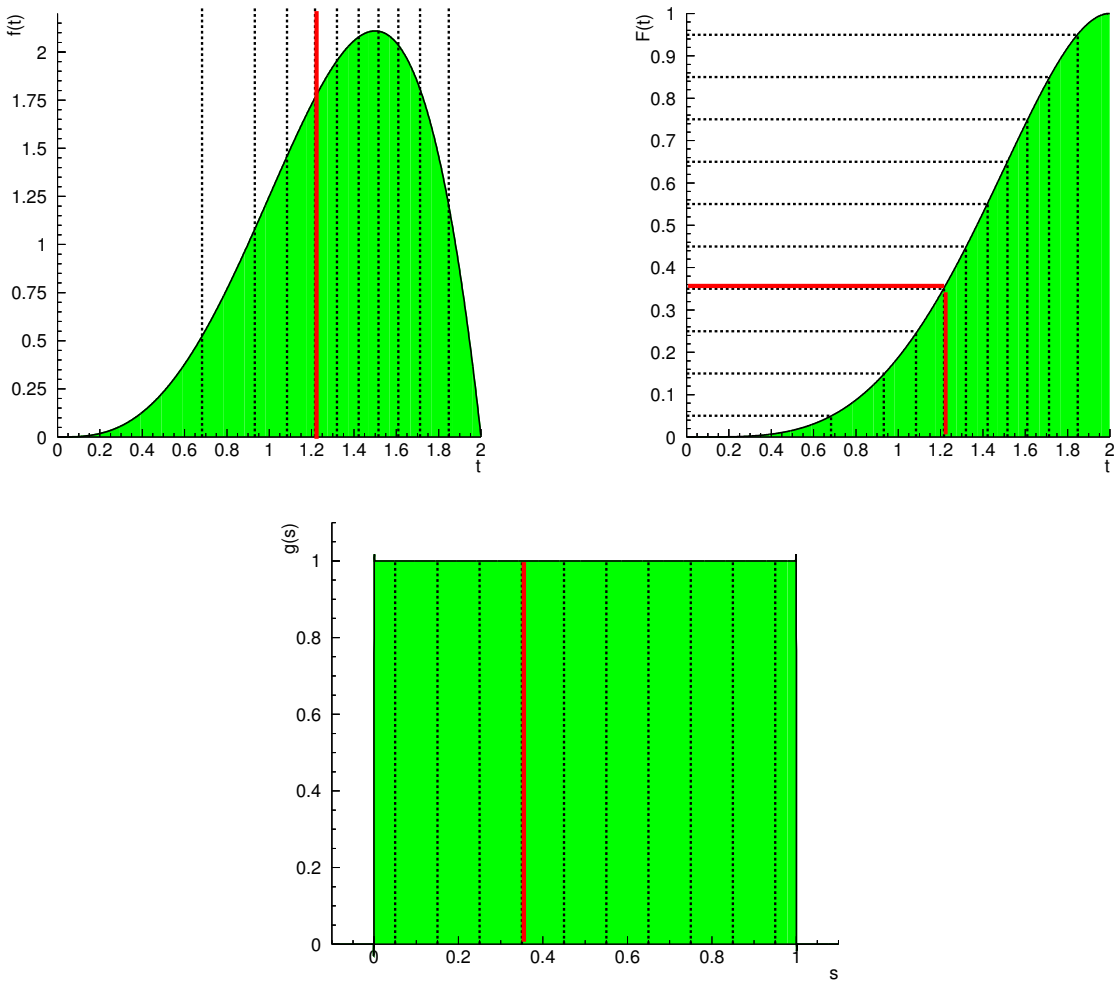


Figure 1.11: Variable transformation during preprocessing.

As shown in the example in figure 1.11, all the distributions $f(t)$ of the input variables (lop left) are transformed to other distributions $g(s)$ during preprocessing by a non-linear variable transformation $F : t \rightarrow s$ (top right in figure 1.11), so that the resulting distribution (signal and background) is flat and within the interval 0 and 1 (bottom in figure 1.11). As a simple example, we use the mass distribution of J/ψ candidates, that have a nominal mass of 3.096 GeV/c^2. The flattened distribution can be seen in figure 1.12. The structure of the original mass distribution can still be seen in this distribution with many signal events and relatively few background events around the nominal mass. This step of flattening the distribution is followed by obtaining the purity in each bin of the new, flattened distribution, and optionally performing a spline fit of the resulting purity distribution, as shown in figure 1.13. The obtained purity distribution

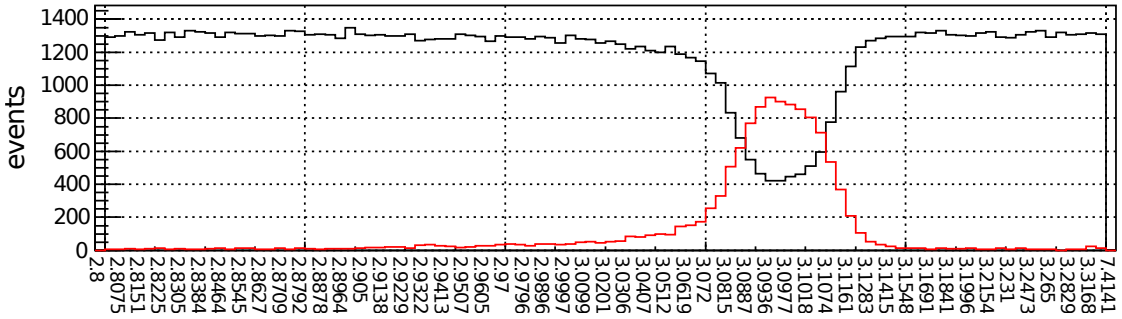


Figure 1.12: Flattened distribution, red: signal, black: background.

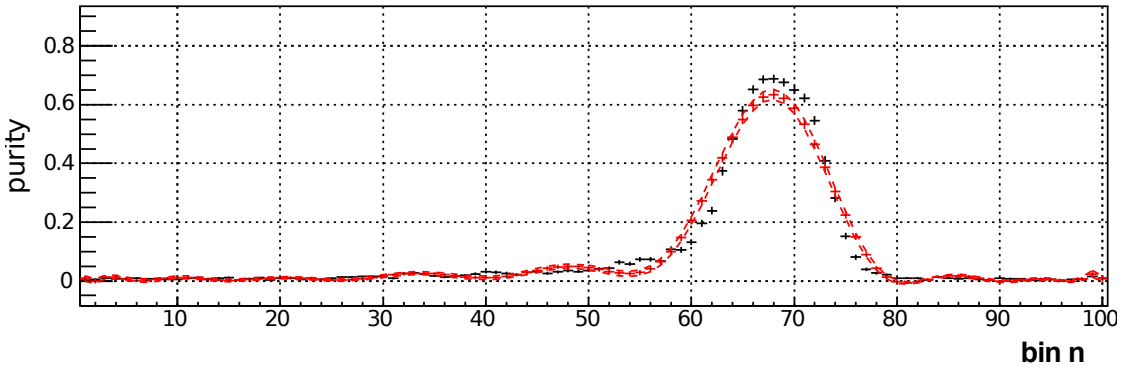


Figure 1.13: Purity distribution, red: spline fit, black: measurement.

is then again transformed, optionally according to the performed fit, to have a mean of 0 and a width of 1. The result of this final transformation can be seen in figure 1.14. To estimate the separation power of this one variable compared to the whole neural network, a purity-efficiency plot is created for each input variable, which can be seen in figure 1.15, where now the black line indicates the purity-efficiency curve for the current variable alone, while the red curve is for all of the input variables combined. Although it would be possible to perform the next step, the training of a neural network, with unprocessed data, there are large benefits in speed and robustness of the algorithm when performing preprocessings.

The Neural Network

The neural networks utilised in the NeuroBayes package share their common topology of a three-layered feed forward neural network. Figure 1.16 shows an exemplary neural network with 5 input nodes, 4 hidden nodes and one output node. It can be shown, that any neural network can be simplified to a three layer network, if there are enough nodes in the hidden layer. Neural networks are based on the working principles of neurons in nature. The input nodes are connected to the nodes in the hidden layer, who produce, when activated, an output to the output node. The output y of the nodes in the hidden

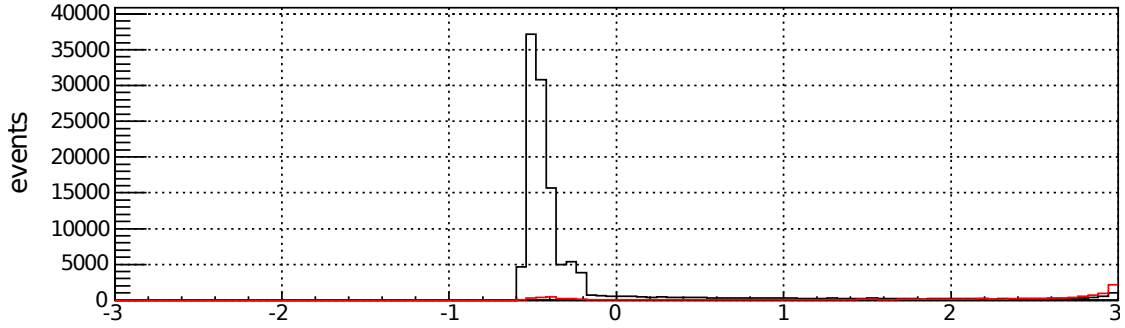


Figure 1.14: Final distribution, red: signal, black: background.

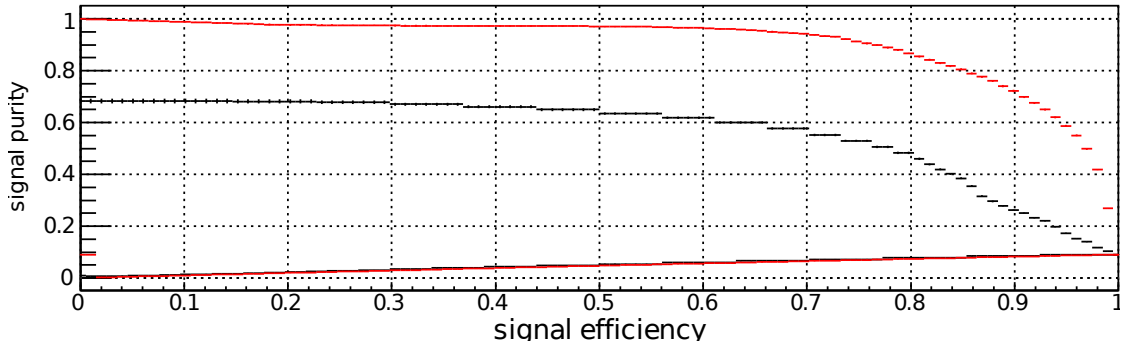


Figure 1.15: Purity efficiency plot, red: current variable only, black: all variable combined.

layer can be expressed as

$$y = S\left(\sum_i w_i x_i\right) \tag{1.23}$$

where x_i are the different inputs of the nodes with their respective weights w_i that can assume any real value. The widths of the connecting lines between the nodes in figure 1.16 represent their individual weights. The function $S(x)$ is called the activation function and is chosen to be a sigmoid function. This function has the virtue to map the whole range of real numbers to the interval between -1 and 1:

$$S: \quad]-\infty, \infty[\quad \longrightarrow \quad]-1, 1[\tag{1.24}$$

The sigmoid function is defined by

$$S(x) = \frac{1}{1 + e^{-cx}} \tag{1.25}$$

and shown in figure 1.17, where we can see that this function has

- two approximately constant regions at relatively large and small values of x,

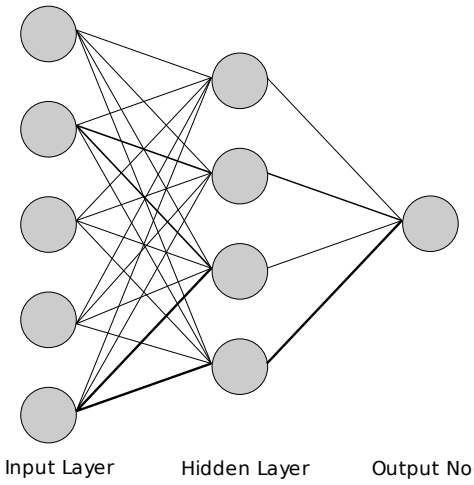


Figure 1.16: Topology of a three-layered feed forward neural network.

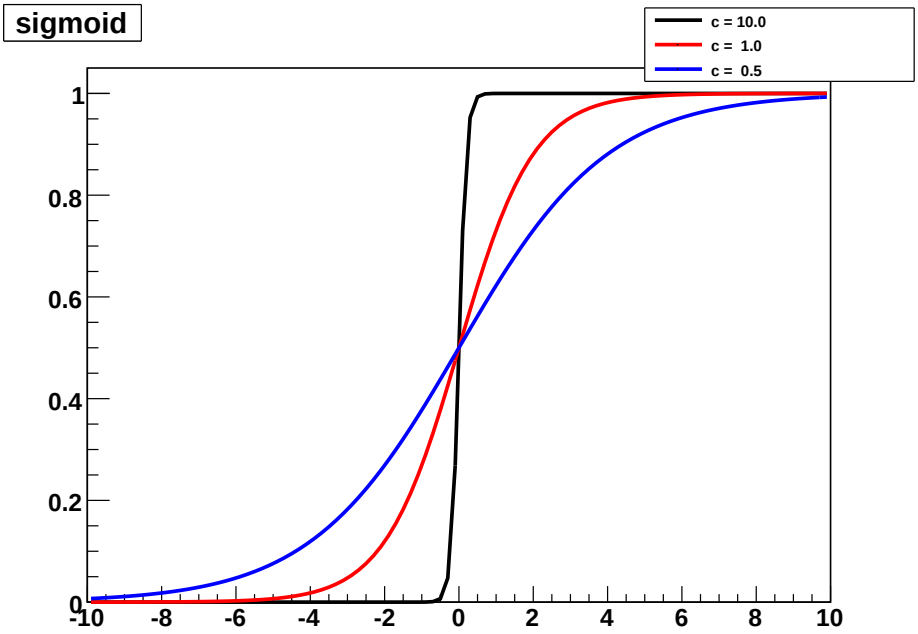


Figure 1.17: The sigmoid function for different values of c.

- a linear region around 0
- an exponential region at negative values near zero and

- a logarithmic region at positive values near zero.

Where exactly these regions are, depends on the choice of the parameter c, which can also be seen in figure 1.17. The shape of the sigmoid function allows not only simple summation of the input variables, as suggested by equation 1.23 but also multiplication, exponential and logarithmic operations. If we now look at the whole three layered network, we can express the output of the node in the output layer o using equation 1.23 twice as:

$$o = S\left(\sum_l w_l^{2\to 3} \cdot S\left(\sum_k w_{kl}^{1\to 2} \cdot x_k\right)\right) \tag{1.26}$$

where $w_{kl}^{1\to 2}$ is the weight of node k of the input layer to node l in the hidden layer and $w_l^{2\to 3}$ is the weight of node l in the hidden layer to the the single output node.

Training

In order to utilise a neural network, first a training has to be performed, the results of which are then applied to characterise the data in question. To perform the necessary training we need a data-set, for which we know the truth. Generally, this can be historical data, but in particle physics, Monte Carlo simulations are often used. The neural network is then trained with a sample of signal and another sample of background events. The target for the neural network is known for each event in the training samples and can be either $t = 1$ for signal or $t = -1$ for background. The process of a neural network training is, after setting the weights described in formula 1.23 initially randomly, the readjustment of these weights in order to describe the known target values as good as possible. In other words, the neural network training is essentially a fit with the weights as free parameters. To measure the quality of this fit, a loss function is used, which has to be minimised for all events i in the sample simultaneously. This can either be a quadratic or an entropy loss function:

$$\chi^2 = w\frac{1}{2}\sum_i (t_i - o_i)^2 \ , \tag{1.27}$$

$$E_D = w\sum_i \log\left(\frac{1}{2}(1 + t_i \cdot o_i + \epsilon)\right) \ , \tag{1.28}$$

where t_i are the known targets ($+1$ or -1) and o_i are the corresponding network outputs. For classification problems, the entropic loss function is generally used. These two loss functions can only be evaluated after a sufficiently large subsample of the available data has been used. In order to achieve an improvement in the prediction power of the network, it is common to perform multiple iterations on the dataset.
This training process by itself contains one common problem of neural networks, however, which is over-training. When the signal (or background) fraction in the training is not large enough and there are many nodes in the input layer, the neural network might be able to learn the signal (or background) events by heart. To avoid this problem,

the concept of decaying weights is used, which means that the weights of all nodes w_i are moved towards 0 after a fixed interval, which can either be one iteration or a fixed number of events. In a figurative way, one could say that the neural network is programmed to forget part of what it has learned to avoid over-training. Additionally the training is not performed on the whole available data-set, but a small fraction of signal and background events are left out of the training and used for validation of the neural network. Over-training would be indicated by exceptionally bad performance on such a validation sample.
While it would be possible, and it sometimes indeed is practical, to train the neural network with the same signal to background ratio as expected on data, it is sometimes more practical and sometimes even only possible using a different ratio. If, for example, the desired signal is very rare in nature, a training would not learn to distinguish the few signal events from the millions of background events, but rather try to learn something from statistical fluctuations that surpass the signal by far and therefore also dominate the loss function. Therefore, a training with a higher signal fraction is the only way, in which such rare signals can be detected.

Performance

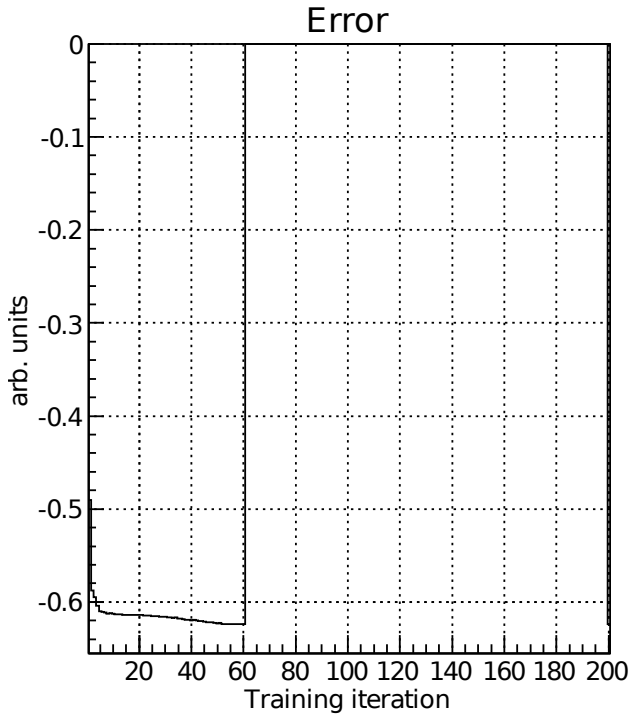


Figure 1.18: The error function during a NeuroBayes training. The training was stopped approximately after the 60th iteration.

After a performed neural network training, there are several plots and figures that provide information about the quality of a network training. First, we have error plots, shown in figure 1.18, that show how large the loss function was during the individual iterations. During a normal training, these value of the error function should decrease towards a minimal value, which then indicates that a local minimum of the loss function is found. As it would not bring much improvement to do further iterations at this point, NeuroBayes can be programmed to stop the training here. This was in fact done in the exemplary plot in figure 1.18 around the 60th iteration.
The quality of the achieved classification can be described by the ability of the network to separate signal and background of a sample. This ability is shown in figure 1.19. We

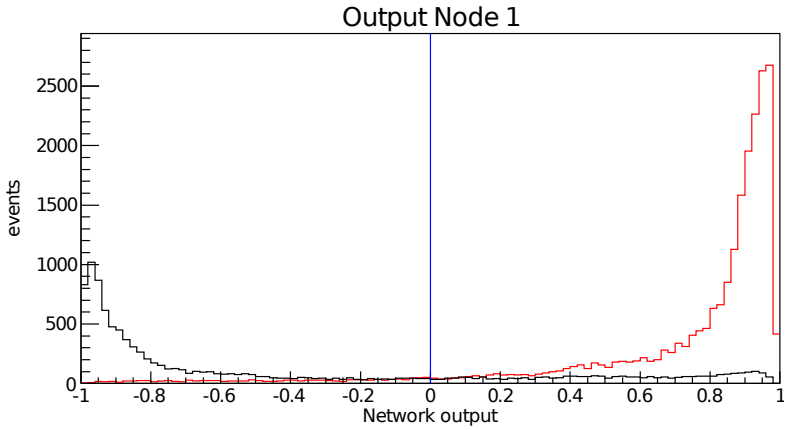


Figure 1.19: The distribution of the NeuroBayes neural network output, red: signal, black: background.

can see a large background peak (black line) at low NeuroBayes outputs and a small signal peak (red line) at high NeuroBayes outputs. This training was done with a signal to background ratio of 1:10, and therefore it is natural, that the signal component is of smaller magnitude than the background. Nevertheless, the good separation is still visible.
The separation, as good as it may look on any plot like in figure 1.19 can never reach 100% signal purity in any realistic situation and, as mentioned in section 1.6, we always have to deal with a compromise between purity and efficiency. The purity-efficiency plot in figure 1.20, automatically generated by NeuroBayes also gives valuable information about the quality of the neural network training. The upper curve shows a usual purity-efficiency curve, that is obtained by calling the candidates above a certain cut "signal" and those below that cut "background". The lower curve is obtained doing just the opposite: Calling the candidates below a certain cut "signal" and those above the cut "background".
Finally, there is one plot that summarises the importance of the given input variables. It is a colour-coded correlation matrix, that assigns each variable a row and a column and additionally introduces the column and row no. 1, which corresponds to the known

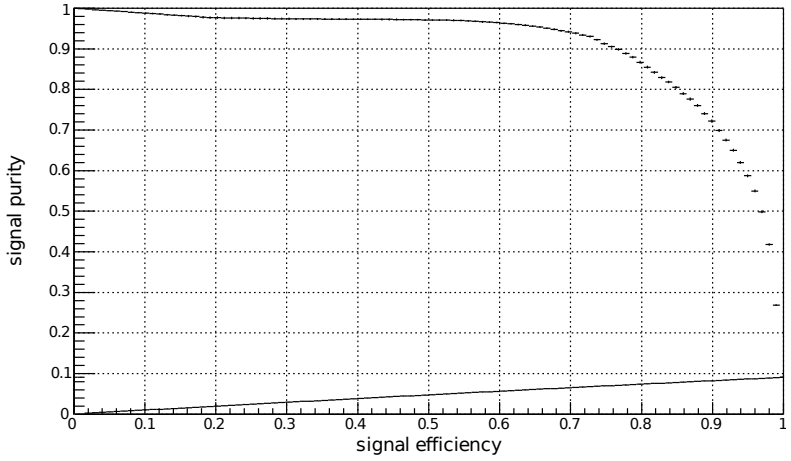


Figure 1.20: The purity-efficiency plot, generated by NeuroBayes.

truth. This correlation matrix shows the correlation of each variable to the truth, but also the correlation of the variables amongst each other. An exemplary correlation matrix is shown in figure 1.21. A strong (anti-)correlation to column no. 1, which is the known truth during a training, indicates the importance of a variable. If this variable is, however, strongly correlated to another variable, only one of these is in fact important for the neural network, as they carry similar information.

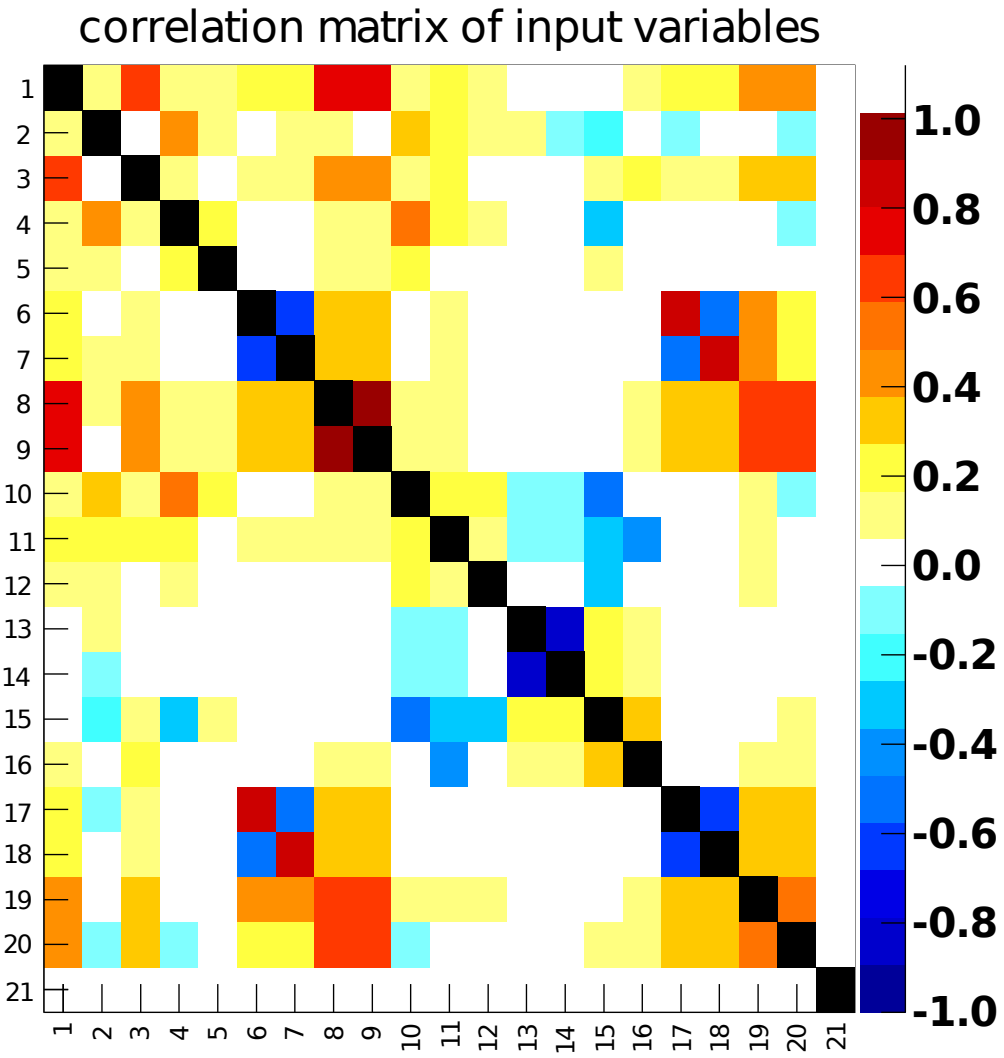


Figure 1.21: Correlation matrix of a NeuroBayes training.

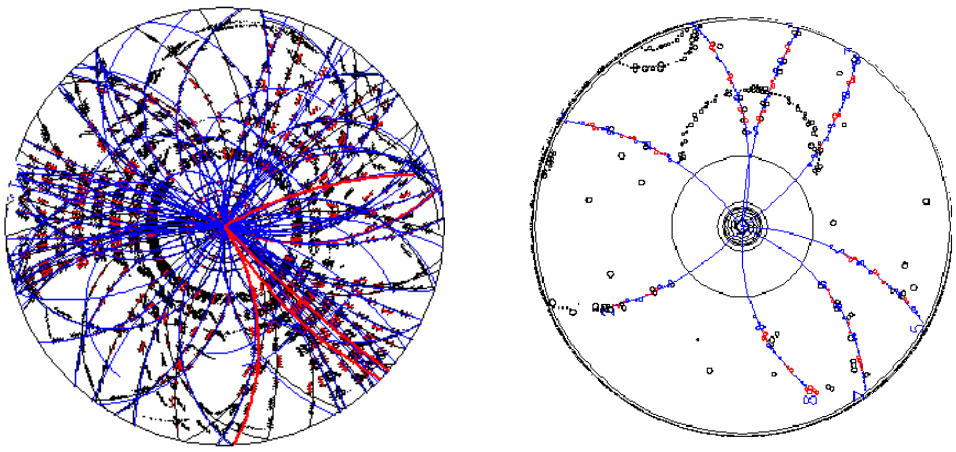
2 Full Reconstruction

2.1 The Basic Idea

2.1.1 Collision Topology

One of the biggest advantages of lepton colliders like the KEK B accelerator or PEP-II compared to hadron accelerators like the Tevatron or LHC is their clean event structure. One property of hadron collisions is that the momentum of a hadron is divided up amongst the sub-particles of the hadron and usually only one sub-particle of each hadron participates in the actual collision. The other constituents of the colliding hadrons do not take any active role in the physical process and are therefore called "spectator quarks". As the momenta carried by the two colliding quarks are not fixed and can indeed vary considerably, the centre-of-mass energy $\sqrt{s}$ of the collision also varies, which results in a vast spectrum of possible processes at different energies. Quite often, the particles produced in the hadron collisions have a lot of phase space left, so that their momenta also vary a lot which leads to fragmentation. This means that there is a rather large amount of tracks and clusters in a typical event at a hadron collider, as shown in figure 2.1(a).

Leptons, on the other hand, do not have any substructure and therefore always take part



(a) CDF event (b) Belle event

Figure 2.1: Different topologies of typical events.

in collisions with their entire momentum, which results in collisions at defined energies without any spectators. The energy of these collision can be adjusted very precisely.

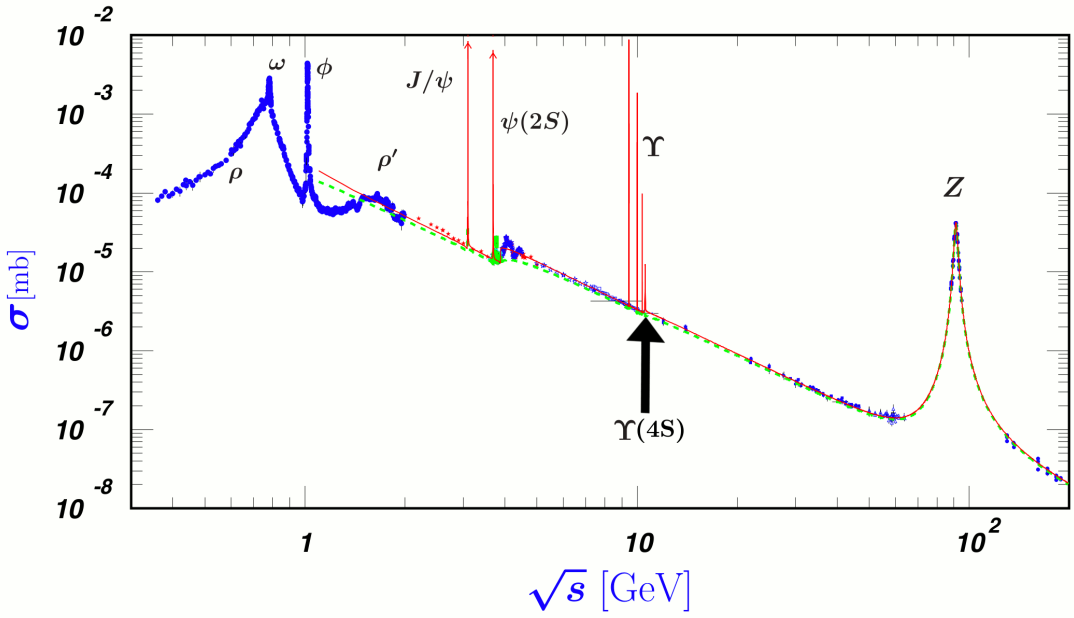


Figure 2.2: Cross section in electron positron collisions.

According to the chosen energy, the production cross section varies largely, as shown in figure 2.2. As the KEK B accelerator and the Belle detector were designed to study B meson decays, the centre of mass energy of the collisions was chosen as $\sqrt{s} = 10.58$ GeV, just above the Υ(4S) resonance, which decays into a B^+B^- or $B^0\bar{B}^0$ pair respectively. The Υ(4S) is indicated in figure 2.2. The two large benefits of choosing this centre of mass energy are:

1. The Υ(4S) resonance decays into a B^+B^- or $B^0\bar{B}^0$ pair respectively without any additionally produced particles.

2. The produced B^+B^- or $B^0\bar{B}^0$ pairs are almost at rest in the centre of mass frame of the Υ(4S):

$$\vec{p}_{B,CMS} \approx 0 \ . \tag{2.1}$$

   These events show a homogeneous event topology with no jet-like structures, as shown in figure 2.1(b).

There are however still events where no Υ(4S), but particles of lower masses are produced. The produced particles in those events bear a lot more momentum in the centre of mass frame and therefore show a jet-like event topology and can be relatively easy rejected. These events are called continuum background and will have to be rejected, as our interest only lies on the events with B meson pair production.

2.1.2 The Process

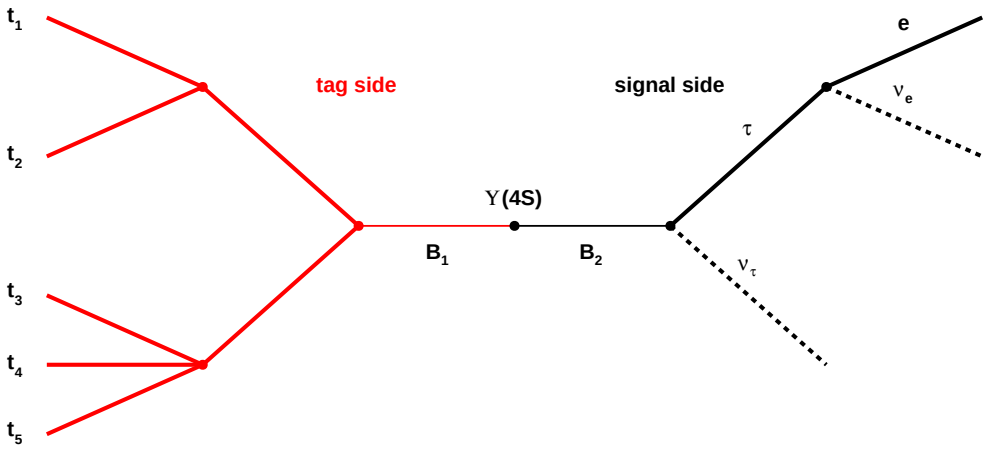


Figure 2.3: Tag and signal side of the full reconstruction.

The process of a full reconstruction is not intended as an analysis in itself, but rather a tool for other analyses. The main goal and also the main difficulty of the full reconstruction is to take any event and try to reconstruct one of the B mesons in many different decay channels. Should one of these attempts succeed, it is possible to assign all the tracks and electromagnetic clusters used during the reconstruction to this one B meson. As it was completely reconstructed, we also know its 4-momentum. We call the fully reconstructed B meson decay the *tag side* and the other B meson is called the *signal side*. The benefit of this procedure lies within the clean event topology at e^+e^- colliders and, ultimately, the defined energy of the KEK B accelerator, as it is now possible to assign all the remaining tracks and electromagnetic clusters within the detector to the other B meson, the *signal side* B. We can be sure that there is no other particle produced by the e^+e^- collision within the detector, as the energy is chosen, as mentioned above, just above the threshold for the production of $\Upsilon(4S)$ resonance. Additionally, as the $\Upsilon(4S)$ decay is a two-body decay, we also know the kinematics of the signal side B meson, without any additional efforts. This procedure might seem rather involved at first glance, but has the benefit that an easy to reconstruct B decay on the tag side can yield otherwise inaccessible information about a hard or impossible to reconstruct B decay on the signal side. A prominent example of a hard to reconstruct decay is a B decay including two or more neutrinos, for example:

$$B^+ \to \tau^+ \nu_\tau \to e^+ \nu_e \nu_\tau \ . \tag{2.2}$$

The topology of this decay is given in figure 2.3. The most important practical difference between this method and most analyses is just the sheer number of used decay channels for the tag side. As there are several hundreds of known B decay channels, and possibly even much more unknown ones in nature, the task of reconstructing one of the two B mesons in the event cannot always succeed. Additionally, most of those decay channels

include other unstable particles, mostly D^* and D mesons, that also decay in a vast spectrum of decay channels, that also have to be reconstructed, of course. As it is technically not sensible to reconstruct multiple hundreds of channels for the B meson alone and then to multiply this number by the used channels for the D^* and D mesons and other unstable particles, two simplifications will have to be applied:

1. There will have to be independent reconstructions for the different types of unstable particles, which will yield sub-samples of D^* and D mesons and other unstable particles and avoid the superfluous task of re-reconstructing for example D mesons for every new B meson decay channel that is used.

2. A good selection of those channels that have large enough branching fractions has to be made, rather than using all known ones.

The hierarchical structure utilising sub-samples and a good selection of decay channels combined should give good results with reasonable consumption of manpower and CPU time. A preliminary list of all decay channels to be used can be found in appendix B.

2.2 Track Identification

During the full reconstruction, it is, as in any other analysis, crucially important to have as few misidentifications as possible. The predecessor of our full reconstruction code comprises a working full reconstruction with a large spectrum of channels and started to include neural networks in the last step, the selection of the B mesons. All the other identifications and selections were done using conventional cuts [Col07]. As it is our hope to increase particle identification performance by using NeuroBayes classification, we decided to use it throughout a new full reconstruction built from scratch.
The full reconstruction starts with charged tracks. The classification criteria provided by combined likelihood ratios like the Belle electron or muon ID or the `atc_pid` likelihood ratios that are available on the `mdst` files were used by the predecessor of the new full reconstruction for hard cuts. In the new full reconstruction, all this particle ID information is now taken as input for neural networks, as described in section 1.6.3. A complete list of variables used for the classification of charged tracks can be found in the appendix A.1 to A.4.. The trainings for these charged tracks were performed on generic Monte Carlo samples. These are samples that contain equal amounts of charged and neutral B meson pairs and additionally the amount of charm and uds background that is expected on data. We hoped to achieve a good general performance on data by training on the general Monte Carlo sample. Matching the reconstructed tracks to the Monte Carlo truth is a simple matter for charged tracks, as the `mdst` files already carry a table that links the tracks to their Monte Carlo counterparts. Several purity-efficiency plots were made to check the quality of the trainings. As a comparison, another curve, which uses standard particle ID cuts [Col07] usually used is plotted. Figures 2.4 to 2.7 show these results.
As we can see from theses plots, all the trainings have been done with a 1:10 signal to background ratio, as with 100% efficiency (which corresponds to selecting simply

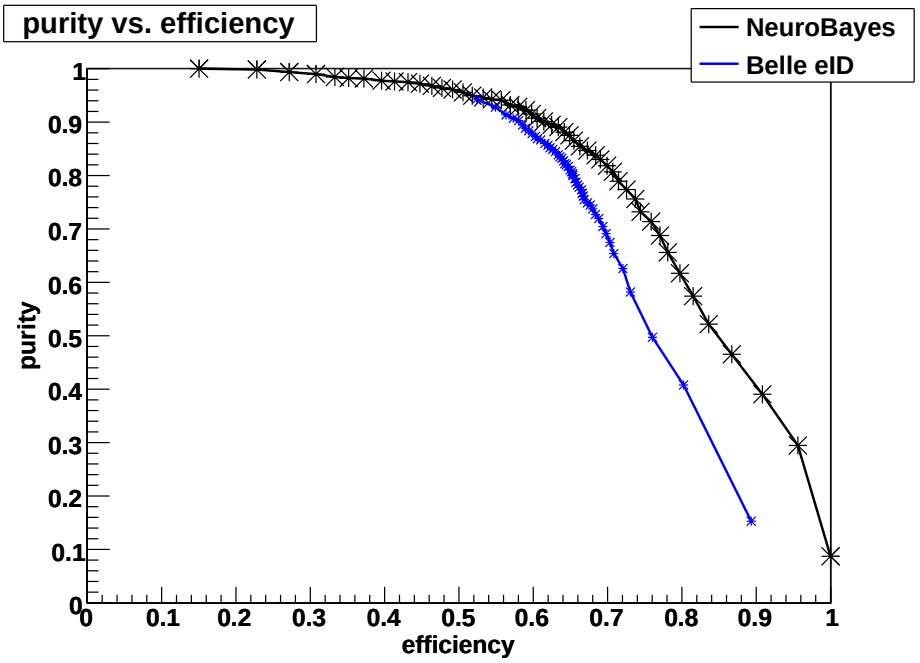


Figure 2.4: Purity-efficiency for electrons.

everything), we achieve a signal purity of 10%. An exception is the $\pi^{\pm}$ training, which was done with a 7:10 signal to background ratio, as this ratio is closer to what we expect on real data. The plots show furthermore, where a NeuroBayes training can make an improvement compared to a combined likelihood ratio like the `atc_pid` and where it can't improve much. In general we don't loose anything, but can improve the identification in a lot of regions on the purity-efficiency plot. The only exception is for the charged kaon identification, where in the area of high efficiencies and very low purities, the cut on the likelihood ratio yields a slightly better performance. The reason for this might be, as the neural network is optimised for the whole range of charged tracks, that there can be drawbacks in boundary areas. As the area, where the drawbacks of the classification of the neural networks occur, is not a usual working point in the purity-efficiency plane and additionally a very small area, the results in figure 2.7 can be accepted.

2.3 Monte Carlo Matching

For the charged tracks, it was easy to perform a NeuroBayes training, as the Monte Carlo information was directly accessible. This is not the case anymore for unstable particles. The default method, also used by the predecessor of this full reconstruction, comprises the counting of stable tracks, matched photon clusters and charged tracks from converted photons, K_S or Λ particles, that were used in the reconstruction of the B meson and the adding up of their energies. By comparison of those two figures with

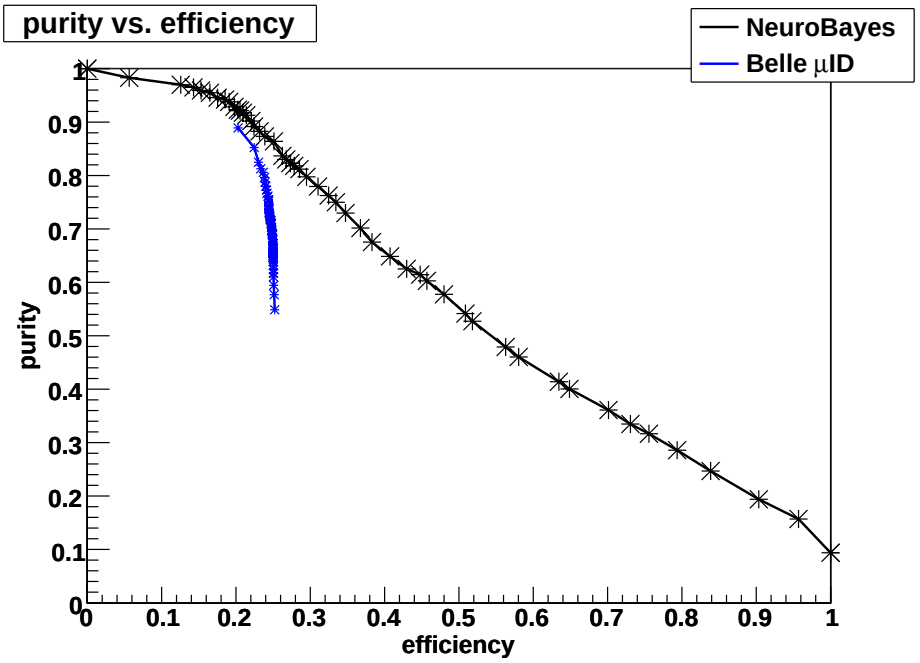


Figure 2.5: Purity-efficiency for muons.

the information on the Monte Carlo tables, it was decided whether a B meson was a *real* one, i.e. matched to Monte Carlo truth, or not. This procedure did not seem sufficient for obtaining clean signal samples for NeuroBayes trainings, as B mesons with mismatched tracks or clusters would still be identified as correctly matched if the energy and number of the used clusters and tracks should happen to match the Monte Carlo tables. One consequence of this was the presence of more than two correctly matched B mesons in a single event, which is energetically impossible. It was decided, that a more precise method had to be developed.
The Monte Carlo matching method that was developed for this full reconstruction is able to handle any kind of stable or unstable particle. It will perform several checks and will quit as soon as either one requirement is not fulfilled, tagging the particle in question as not matched (`mcinfo=-1`). Should the procedure terminate successfully, then the particle in question will be tagged as matched (`mcinfo=1`). The newly developed procedure of Monte Carlo matching can be divided into 3 stages:

1. **Stable tracks and electromagnetic clusters** The algorithm first goes recursively down the decay chain until it reaches the stable tracks and clusters, for which Monte Carlo information is available. If all of these particles have been assigned the correct particle ID during reconstruction and if there are no other tracks or clusters on the Monte Carlo table that should have been reconstructed for the B meson in question, in short, if all the stable daughters are found, we continue to the next stage.

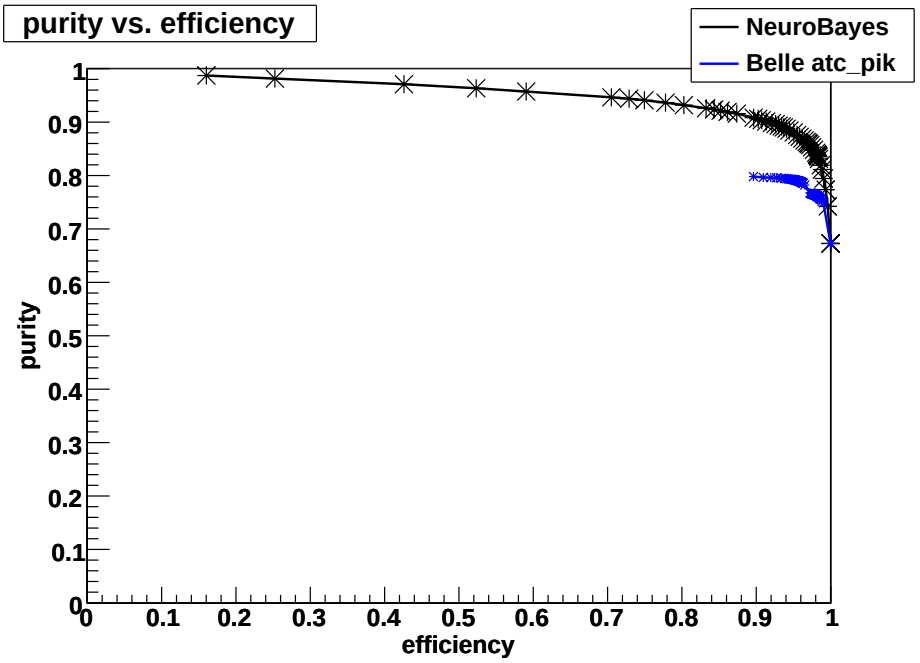


Figure 2.6: Purity-efficiency for pions.

2. **Relations** The algorithm now goes up one step the decay chain again. The new object in question is the "mother" of the particle before. The algorithm now checks, if all of the necessary children for an unstable particle have been assigned to the current particle correctly.
An important question arises when looking at this procedure: If the true decay went, according to the Monte Carlo tables, via a resonance, but the resonance was not reconstructed explicitly, should the particle be treated as correctly reconstructed or not, given an otherwise correct reconstruction? An example of this would be

$$\begin{aligned} &\bar{D}^0 \to K^+ a_1(1260)^- \to K^+\pi^-\rho(770)^0 \quad \to K^+\pi^-\pi^+\pi^- \qquad \text{(MC truth)} \\ &\bar{D}^0 \qquad\qquad\qquad\qquad\qquad\qquad\quad \to K^+\pi^-\pi^+\pi^- \quad \text{(Reconstruction)} \end{aligned} \tag{2.3}$$

The question is now, if the $\bar{D}^0$ can still be regarded as correctly reconstructed or not. As the two resonances missing in the reconstructed decay ($a_1(1260)^-$ and $\rho(770)^0$) are extremely short-lived and the final state is identical, the two decays in equation 2.3 are, by virtue of quantum mechanics, indistinguishable and therefore the reconstructed channel should be treated, given that all other requirements are fulfilled, as correctly reconstructed. We can achieve this by allowing particles in the Monte Carlo tables not only to be assigned to their mother, but also to the next one in the decay channel under certain requirements. We call this procedure "jumping over" mothers and we have the choice to allow this either when a particle

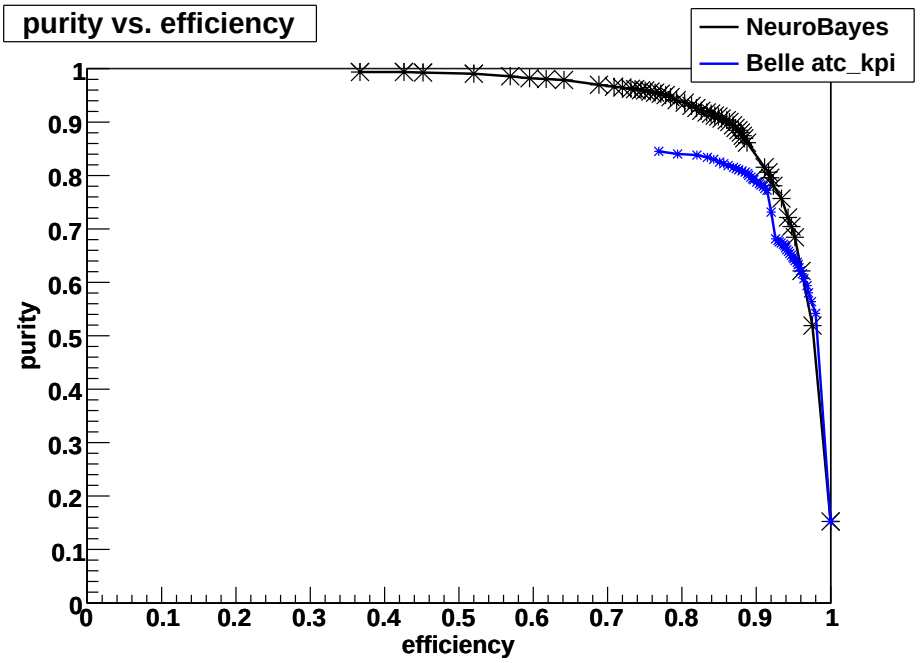


Figure 2.7: Purity-efficiency for $K^{\pm}$.

has lifetime $\tau = 0$ on the Monte Carlo tables, or we decide, based on a hand coded list, which particles may be jumped over. We decided to do the latter.

3. **Final Checks** If we reach this point, we have found all of the stable children of an unstable particle and were able to verify their reconstructed relations on the Monte Carlo tables. Some final checks are performed to avoid possible ambiguities: We check for duplicate children, i.e. two particles in reconstruction, that correspond to the very same particle on the Monte Carlo tables and we check, if the number of stable children in reconstruction and on the Monte Carlo tables are identical. Should any of these checks fail, the algorithm quits and the particle will be tagged as not matched.

If all those requirements have been fulfilled and all checks have been passed, Monte Carlo truth information is assigned to the current particle. This procedure is repeated until the original unstable particle, that was originally passed to the algorithm, is reached.
This new algorithm was tested on the predecessor of the full reconstruction and it worked as intended for all kinds of unstable and stable particles and it got rid of all the ambiguities of its predecessor.

2.4 Outlook

After the classification of the charged tracks, the next logical step would be the combination of those tracks into unstable particles. At this point, the effort in the full reconstruction was adjourned in favour of newsworthy analysis. This analysis, described in detail in chapter 3 was built upon the foundation of the full reconstruction described in this chapter and therefore enabled us to test and improve the concepts and technicalities of our code. Work on the full reconstruction has continued for a while now and experiences from the the analysis in chapter 3 could be incorporated.

3 Search for the $Y(4140)$

Quarks do not occur as free particles, but only in bound states of two or three quarks, which we call mesons or baryons respectively. Another possibility, that has not been observed yet is a bound state of four quarks, a so-called tetraquark. The following chapter describes the search for a resonance that might be a candidate for such a tetraquark. In addition to the analysis described below, another analysis on the same topic was done within the Belle collaboration in parallel. It served this analysis as a cross-check and vice versa.

3.1 CDF Evidence

Recently, the CDF collaboration has found evidence for a narrow near-threshold structure in the $J/\psi\ \phi$ mass spectrum [T. 09] in the decay

$$B^{\pm} \to J/\psi\ \phi\ K^{\pm}\ . \tag{3.1}$$

The Feynman diagram for the positively charged B meson decay is shown in figure 3.1. The structure's mass and width were measured to be

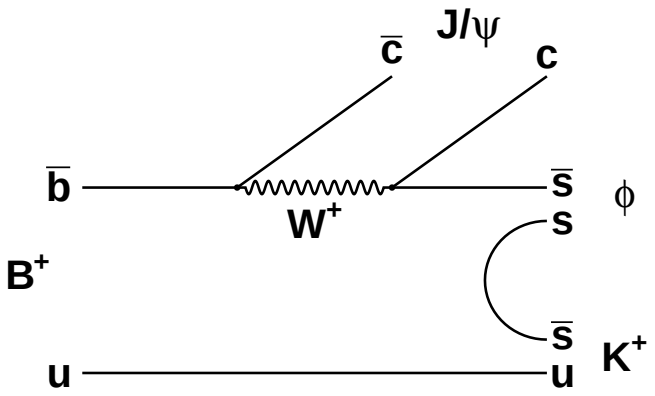


Figure 3.1: Feynman diagram for $B^+ \to J/\psi\ \phi K^+$.

$$m = 4143 \pm 2.9(\text{stat}) \pm 1.2(\text{sys})\ \text{MeV}/c^2\ , \tag{3.2}$$

$$\Gamma = 11.7^{+8.3}_{-5.0}(\text{stat}) \pm 3.7(\text{sys})\ \text{MeV}/c^2\ . \tag{3.3}$$

These measurements were performed in a data sample that corresponds to an integrated luminosity of 2.7 fb^{-1} collected by the CDF II detector. A selection of 75 ± 10 B^+ events yielded a signal of 14 ± 5 events with a significance of 3.8 standard deviations. The $Y(4140)$ structure does not fit into conventional expectation for a charmonium state[1], as it is energetically well above the threshold for open charm decays, but decays into $J/\psi\ \phi$. In analogy with the $Y(3930)$, which lies just above the $J/\psi\ \omega$ threshold and decays into J/ψ and ω, the newly found resonance was named $Y(4140)$. This section will describe the attempts made to confirm or disprove the CDF II evidence with Belle data.

3.2 Reconstruction Outline

The following analysis was built on top of the foundation of the full reconstruction code. The improved classification of charged tracks by the application of neural networks that was described in section 2.2 is available for the $Y(4140)$ analysis. However, only the electron, muon and $K^\pm$ classification networks were needed during this analysis.
To obtain a $Y(4140)$ signal sample, we have to reconstruct the decay channel $B^\pm \to J/\psi\ \phi\ K^\pm$. So we first need samples of J/ψ and ϕ. We reconstruct the J/ψ in two different channels, $J/\psi \to e^+\ e^-$ and $J/\psi \to \mu^+\ \mu^-$ and in order to obtain a ϕ sample, we reconstruct $\phi \to K^+\ K^-$. Using this strategy, we are not able to reconstruct every last J/ψ or ϕ, as these unstable particles also decay into other channels than the ones we reconstruct, but reconstruction of the other decay channels is rather involved and can only gain very little compared to the necessary effort. The decay channels of the $B^\pm$ mesons that were reconstructed for this analysis and their respective branching ratios can be found in table 3.1.

Channel			Branching ratio
$B^\pm$	$\to$	$J/\psi\ \phi\ K^\pm$	$(5.2 \pm 1.7) \times 10^{-5}$
J/ψ	$\to$	$e^+\ e^-$	$(5.94 \pm 0.06)\%$
J/ψ	$\to$	$\mu^+\ \mu^-$	$(5.93 \pm 0.06)\%$
ϕ	$\to$	$K^+\ K^-$	$(49.2 \pm 0.6)\%$

Table 3.1: Decay channels and branching ratios for the $Y(4140)$ search [eaPDGon].

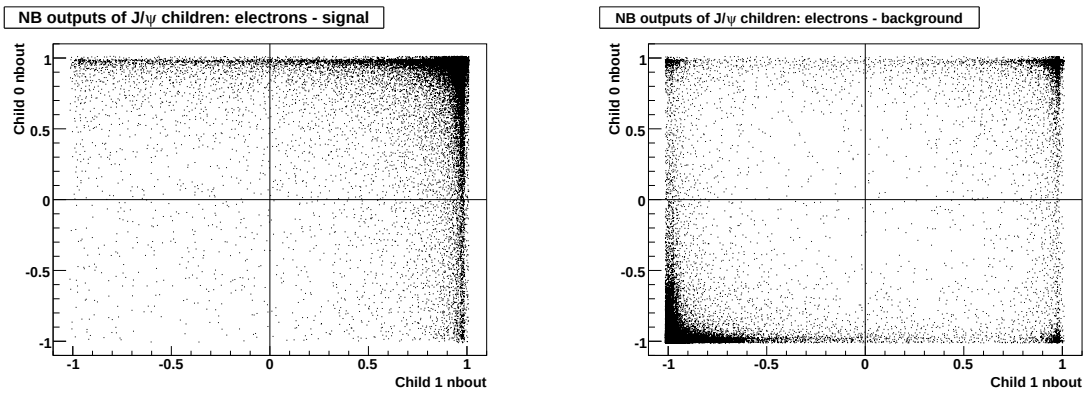
3.3 J/ψ Reconstruction

3.3.1 Combinations

The leptons used for the J/ψ reconstruction were classified as described in section 2.2 using neural networks, but no cuts have been applied so far, neither on the usual likelihood ratios provided by Belle (`electron_ID` or `muon_ID`) nor on the output of the neural

[1] A bound state of a c and $\bar{\text{c}}$

networks. It is, however, not reasonable to combine every lepton candidate with every other one, as this would basically mean combining all charged tracks. On the other hand, we found that performing a hard cut on all of the lepton candidates reduces signal efficiency too strongly, so a different approach was chosen: For the reconstruction, at least one good lepton is required. A lepton is considered good when its NeuroBayes output exceeds a certain level. For the electrons and muons used for the J/ψ reconstruction, the cut was chosen at a NeuroBayes output of 0.0. Each lepton from this good sample is then combined with any other lepton. So no lepton is actually cut, just some combinations are not made. This gives us an efficiency of 97.0%, with a background suppression of 15.9. To achieve the same background suppression with a cut on the NeuroBayes output of both leptons, the efficiency drops to 90%. Figures 3.2 and 3.3 provide visualisations of the requirement of at least one good lepton, for electrons and muons respectively. Requiring at least one good lepton means leaving out the lower left corner of the plots in figures 3.2 and 3.3.



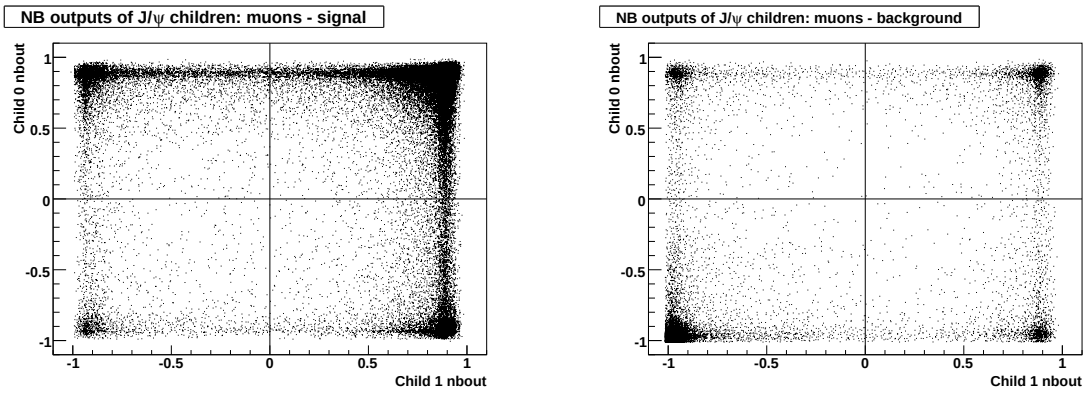
(a) Signal on generic Monte Carlo

(b) Background on generic Monte Carlo

Figure 3.2: NeuroBayes outputs for the electrons used for reconstruction of J/ψ.

3.3.2 Photon recovery

When we only deal with the J/ψs that decay into electrons, the issue of final state radiation comes to mind. This phenomenon comprises the emission of a photon in the process of the J/ψ decay. Another mechanism by which photons are emitted by the electrons from the J/ψ decay is bremsstrahlung. This process always occurs when charged particles are deflected by the atomic nuclei of the medium through which they travel. A consequence of both effects is missing energy of the reconstructed J/ψ, carried away by the photon. This may result in actual J/ψ particles that have a too low reconstructed mass and may therefore wrongfully not be selected as signal. The missing photons are, however, detected by the electromagnetic calorimeter and can be assigned to the electrons they were emitted by. They are, in a manner of speaking, recovered.



(a) Signal on generic Monte Carlo

(b) Background on generic Monte Carlo

Figure 3.3: NeuroBayes outputs for the muons used for reconstruction of J/ψ.

The requirement we chose for this recovery is an isolated[2] cluster in the electromagnetic calorimeter that lies within a narrow angle $\alpha < 5°$ of the initial direction of the electron. This way, we recover photons from final state radiation and also some photons from bremsstrahlung. But as bremsstrahlung can occur along the entire track of the electron, our photon recovery cannot recover all of them. Figure 3.4 shows the effects of the described recovery procedure. We can see clustering at the nominal J/ψ mass of 3.096 GeV after photon recovery (y-axis). We take the clustering as an indication, that the added photons were indeed coming from final state radiation of decaying J/ψ particles.

3.3.3 Classification

The classification of the J/ψ candidates was done by training a neural network. A detailed list of its input variables and their importance can be found in appendix A.5. The training was performed on a generic Monte Carlo sample with a signal to background ratio of 1:10 and $\sim$ 29000 signal events. The resulting purity efficiency plot is shown in figure 3.5.

3.4 ϕ Reconstruction

3.4.1 Combinations

The ϕ was reconstructed in the channel

$$\phi \to K^+K^- \ . \tag{3.4}$$

In order to get the necessary $K^\pm$ sample, there was again no cut performed on one of the usual likelihood ratios (e.g. `atc_kpi`), but the classification of a neural network

[2] An *isolated* cluster is a cluster in the electromagnetic calorimeter, that cannot be matched to a charged track in the drift chamber.

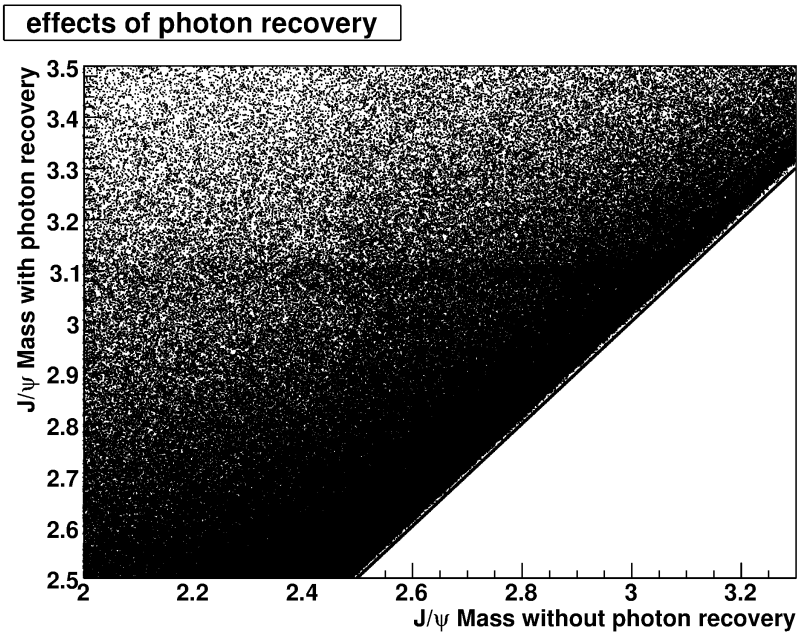


Figure 3.4: Effects of photon recovery.

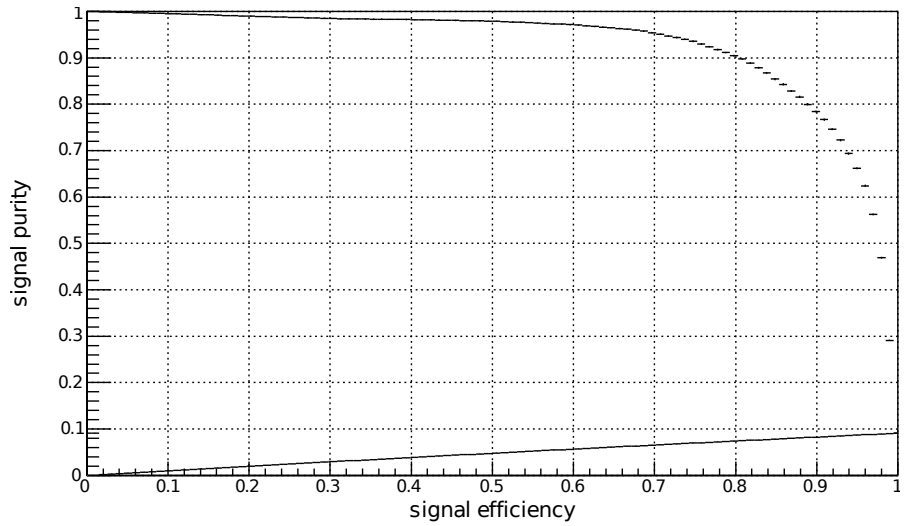
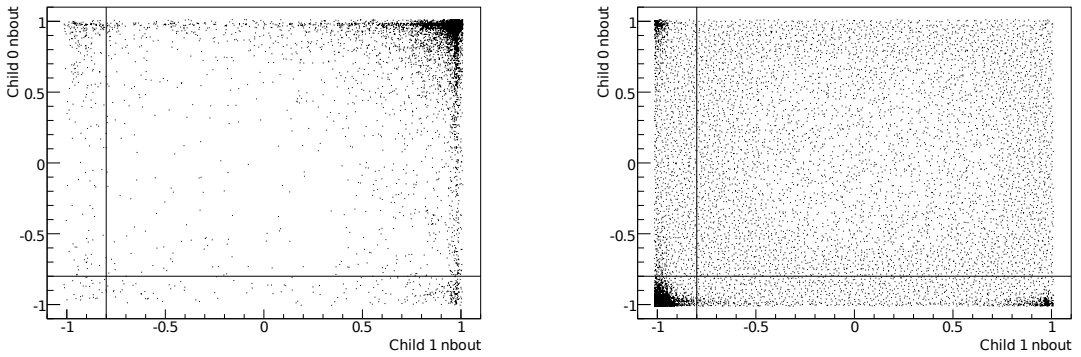


Figure 3.5: Purity efficiency plot for the J/ψ neural network.

was used. Using the output of this network, the same selection criterion as for the J/ψ candidates was used: We required at least one good $K^{\pm}$ for the reconstruction of the ϕ. The NeuroBayes cut that was chosen for a $K^{\pm}$ to be part of the good sample was -0.8. Figure 3.6 visualises the combinatorics.



(a) signal on generic Monte Carlo

(b) Background on generic Monte Carlo

Figure 3.6: NeuroBayes outputs for the charged Kaons used for reconstruction of ϕ.

3.4.2 Classification

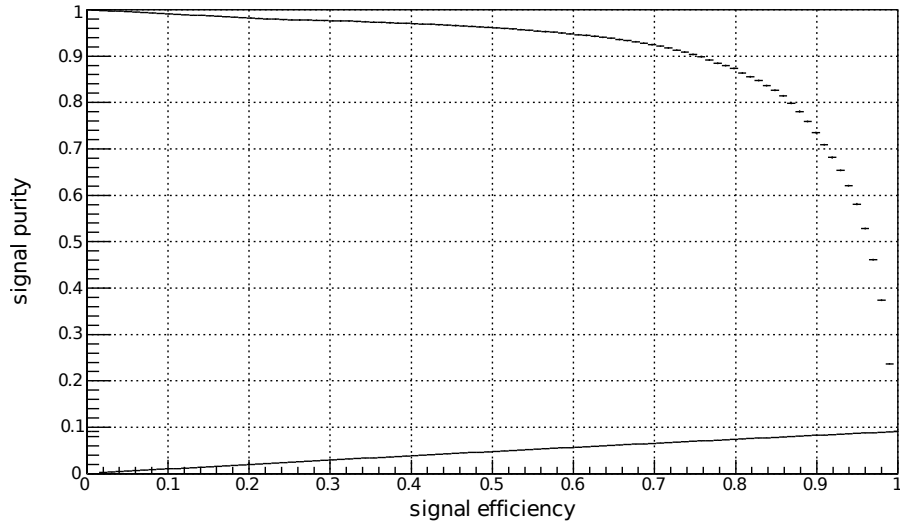


Figure 3.7: Purity efficiency plot for the ϕ neural network.

The classification of the ϕ candidates was done by training a neural network. A detailed list of its input variables and their importance can be found in appendix A.6. The training was performed on a generic Monte Carlo sample with a signal to background ratio of 1:10 and $\sim$ 30000 signal events. The resulting purity efficiency plot is shown in figure 3.7.

3.5 $B^{\pm}$ Reconstruction

3.5.1 Combinations

Before the $B^{\pm}$ mesons are finally reconstructed, very soft precuts are applied on the NeuroBayes output of J/ψ, ϕ and $K^{\pm}$. These are the first and only real cuts to be applied during this reconstruction. The values of the precuts and their individual and combined background reduction factors and efficiencies are given in table 3.2. The pre-

Particle	NeuroBayes cut	Background reduction	Efficiency
J/ψ	-0.980	5.7	99.0%
ϕ	-0.990	8.3	97.6%
$K^{\pm}$	-0.990	1.7	99.9%
All combined		77.6	96.0%

Table 3.2: Precuts for the $B^{\pm}$ reconstruction.

cuts are performed in order to cut away the most obvious background with a background reduction factor of 77.6 and at the same time an efficiency of 96.0%. The benefit of these cuts is, that they enables the neural network to learn the more subtle criteria of the classification. Without the precuts, the neural network would learn just the trivial course of action, which is rejecting the huge amount of obvious background, identified by very low NeuroBayes outputs.

3.5.2 Classification

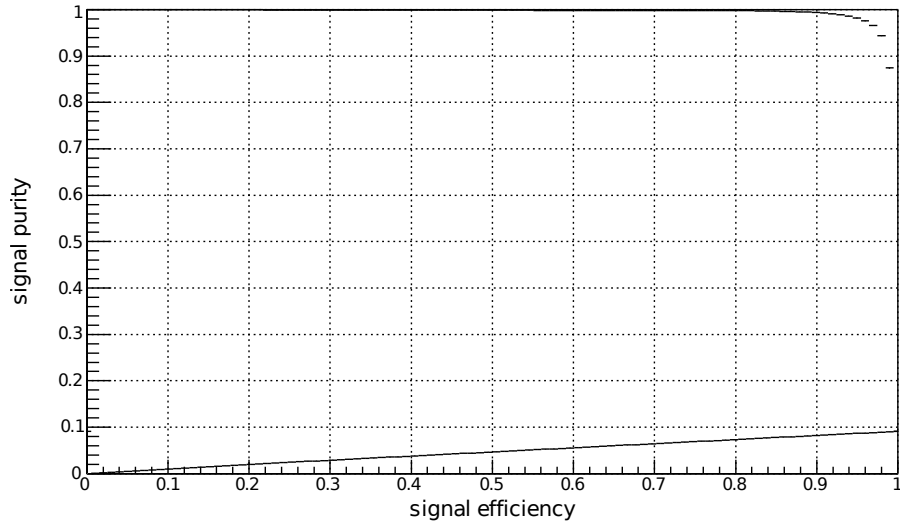


Figure 3.8: Purity efficiency plot for the $B^{\pm}$ neural network.

The classification of the $B^{\pm}$ candidates was done by training a neural network. A detailed

list of its input variables and their importance can be found in appendix A.7. The signal part of the training sample was taken from a $B^{\pm} \to J/\psi\ \phi\ K^{\pm}$ phase space signal Monte Carlo and only the correctly reconstructed events were used. The background sample was taken from a generic Monte Carlo, where only the incorrectly matched events were used. A signal to background ratio of 1:10 was chosen and ~ 36000 signal events were used for the training. The resulting purity efficiency plot is shown in figure 3.7. This purity efficiency plot gives the impression of an extremely good separation, which is true due to the fact that information about the mass of the $B^{\pm}$ was included in the network training. Specifically, the variable M_{BC} was included, which can be written as:

$$M_{BC} = \frac{1}{c^2}\sqrt{E^{*2}_{\mathrm{Beam}} - c^2 p_B^2} = \frac{1}{c^2}\sqrt{E^{*2}_{\mathrm{Beam}} - c^2\left(\sum_i p_i\right)^2}\ , \tag{3.5}$$

where E^{*}_{Beam} is the nominal beam energy, p_i are the momenta of the reconstructed daughters of the B meson and i runs over all of the reconstructed daughters. The inclusion of M_{BC} in the network training prevents us, however, from using the M_{BC} distribution as a quality criterion, as it is common practice. We use instead the ΔE distribution. ΔE is defined as:

$$\Delta E = E_B - E^{*}_{\mathrm{Beam}} = \sum_i \left(\sqrt{c^2 p_i^2 + c^4 m_i^2}\right) - E^{*}_{\mathrm{Beam}} \tag{3.6}$$

where, again E^{*}_{Beam} is the nominal beam energy, p_i are the momenta of the reconstructed daughters of the B meson and i runs over all of the reconstructed daughters.

3.6 $B^{\pm}$ Analysis

In order to check the quality of our reconstruction and selection process, and to obtain a final $B^{\pm}$ sample, ΔE plots were produced. In order to obtain the signal shape, the $B^{\pm} \to J/\psi\ \phi\ K^{\pm}$ phase space signal Monte Carlo was used. An exemplary plot with a NeuroBayes cut of 0.81 can be seen in figure 3.9. The asymmetry we can observe in this distribution was investigated and we found that it comes from the way we did the combination of the leptons for the reconstruction of the J/ψ candidates (see section 3.3.1). The shape of the distribution was fitted with a double Breit-Wigner distribution. There is no physical reason for choosing this function, but it was chosen, because it phenomenologically fits the shape in figure 3.9 so well. The shape of this double Breit-Wigner was then fixed and fitted to data, assuming a flat background distribution. As a data sample, the "Psiskim" of experiments 7 to 65 was used, which corresponds to $771 \times 10^6\ B\bar{B}$ pairs. The fitting procedure yielded approximate numbers for expected signal and background events in the $B^{\pm}$ sample. The exemplary plot obtained for a NeuroBayes cut of 0.94 is shown in figure 3.10. In section 3.6.1, it will be shown, that this cut is a good choice for the analysis and it was indeed used. We define the signal region as the area -40 MeV $< \Delta E < 40$ MeV, as this cuts away almost no signal, but a large amount of background. We get 285 ± 18 signal events with a background level of 78 ± 4 events.

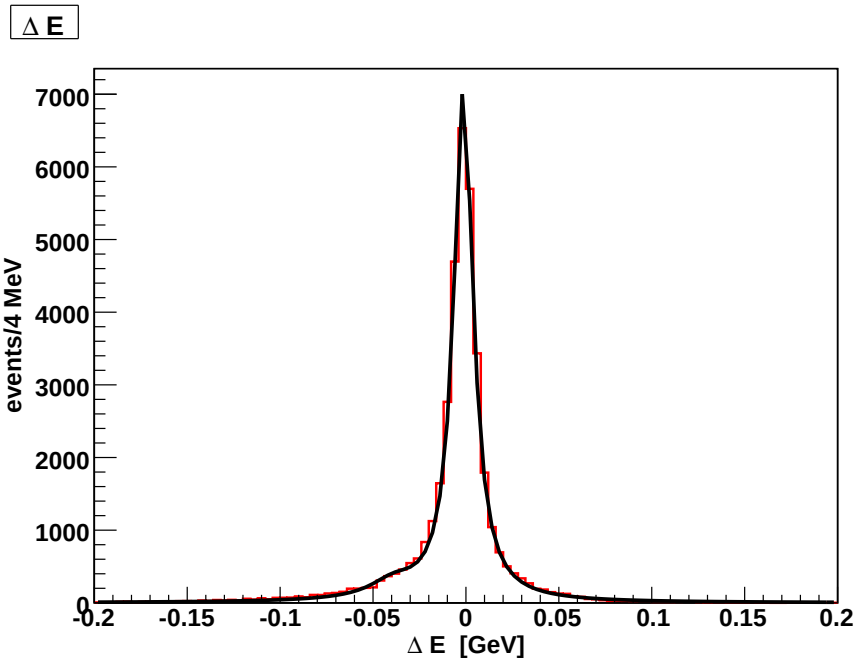


Figure 3.9: ΔE distribution for $B^{\pm} \rightarrow J/\psi\ \phi\ K^{\pm}$ signal Monte Carlo.

3.6.1 Expected Significance

In order to confirm or disprove the observation made by the CDF collaboration (see section 3.1), we aim for a high significance in the mass region of the $J/\psi\ \phi$ mass spectrum, where we expect the $Y(4140)$ signal. A $Y(4140)$ signal Monte Carlo was produced, using the mass and width, measured by the CDF collaboration (see equation 3.2). This Monte Carlo sample was then used to obtain the shape of the $J/\psi\ \phi$ mass spectrum under the hypothesis that the $Y(4140)$ exists. A plot of all the signal events can be seen in figure 3.11. As practically all of the signal lies within the interval $4.1\ \text{GeV}/c^2 < M(J/\psi\ \phi) < 4.2\ \text{GeV}/c^2$, we aim to maximise the expected significance in this interval of the $J/\psi\ \phi$ mass spectrum. In order to do so, as always, we have to find a compromise between good purity and good efficiency and choose the NeuroBayes cut in such a manner, that the significance is maximised. To calculate the expected significance, we first need the expected number of signal and background events.

Expected Number of Signal Events

In order to get an estimate for the number of expected $Y(4140)$ signal events, we need to have an estimate of the branching ratio, the efficiency of our reconstruction and selection process and the number of $B\bar{B}$ events:

$$N_{Y(4140)} = \text{N}(B\bar{B}) \cdot \text{BR for } Y(4140) \cdot \text{efficiency} \cdot J/\psi \text{ and } \phi \text{ BR} \ . \tag{3.7}$$

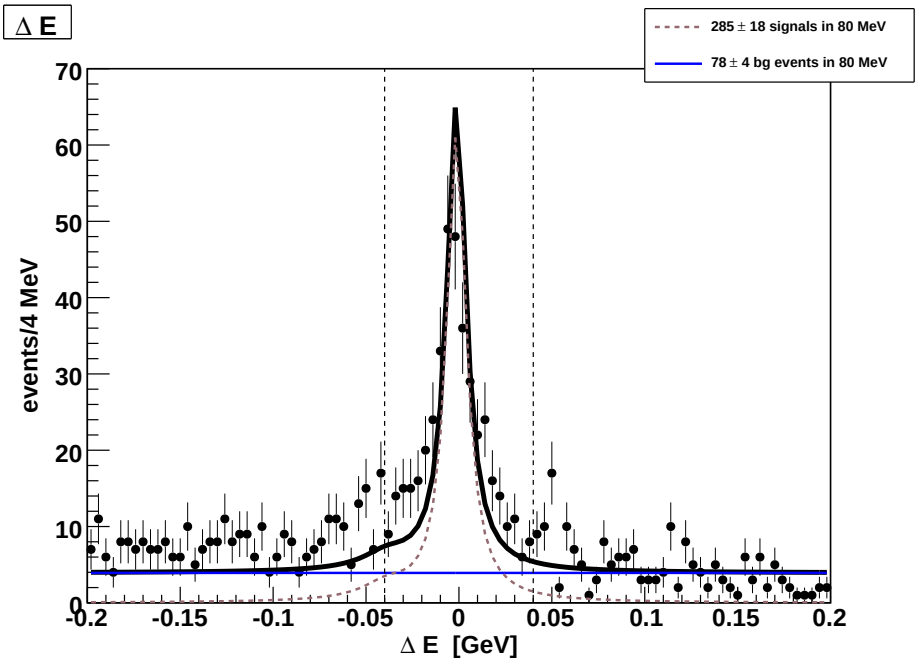


Figure 3.10: ΔE distribution for $B^{\pm} \to J/\psi\ \phi\ K^{\pm}$ data. The dotted lines visualise the 80 MeV area that is selected.

The number of used $B\bar{B}$ events for the experiments 7 - 65, that were used for this analysis is $\sim (771 \pm 10) \times 10^6$. The CDF collaboration estimated the branching ratio of the $Y(4140)$ with $(9.0 \pm 4.5) \times 10^{-6}$. This branching ratio has a relatively big error and we will therefore obtain quite imprecise expectations. We must furthermore take into account, that we do not reconstruct all of the J/ψ and ϕ particles, but only

$$J/\psi \to e^+\ e^- \quad , \tag{3.8}$$

$$J/\psi \to \mu^+\ \mu^- \quad \text{and} \tag{3.9}$$

$$\phi \to K^+\ K^- \quad . \tag{3.10}$$

This again reduces the expected number of signal events. A factor of the summed up branching ratios of the J/ψ and ϕ particles respectively (see table 3.1) will account for this fact. The last number that is missing in equation 3.7 is the efficiency of the reconstruction and selection process. This efficiency depends on the NeuroBayes cut that we apply during the selection. This dependence is shown in figure 3.12. It is not surprising, that the highest efficiency is achieved with the lowest NeuroBayes cut, as there is nothing cut at all and therefore the entire reconstructed signal remains in the sample. We can now calculate the number of expected $Y(4140)$ signal events using

$$N_{Y(4140)} \approx \underbrace{771 \times 10^6}_{\#B\bar{B}} \cdot \underbrace{(9.0 \pm 4.5) \times 10^{-6}}_{\text{CDF-BR for } Y(4140)} \cdot \text{efficiency} \cdot \underbrace{0.1187 \cdot 0.492}_{J/\psi \text{ and } \phi \text{ BR}} \ . \tag{3.11}$$

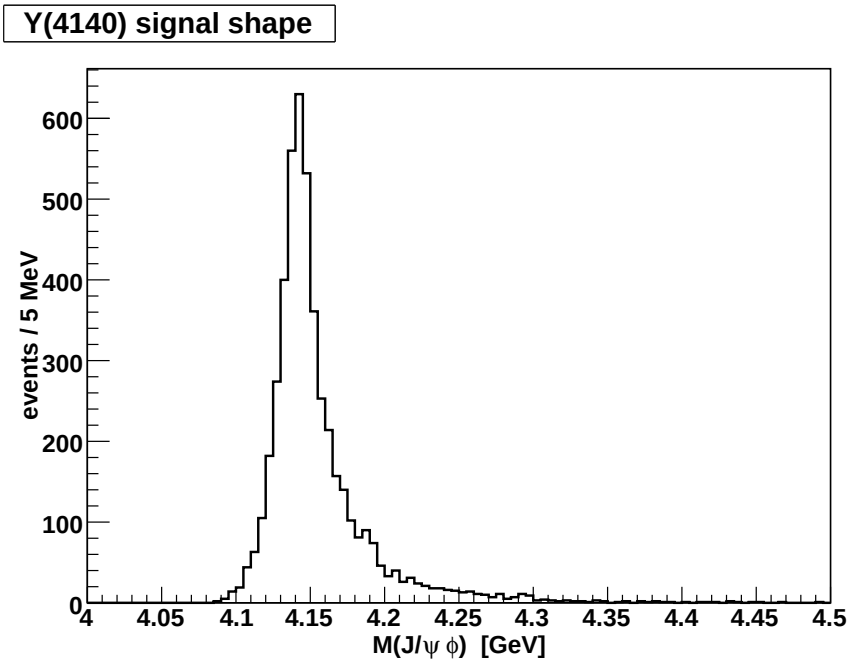


Figure 3.11: Signal shape for $Y(4140)$ signal Monte Carlo.

We disregarded the errors on the number of $B\bar{B}$ events and the branching ratios of the J/ψ and ϕ particles, as these errors are tiny compared to the error of the $Y(4140)$ branching ratio. The resulting plot of the expected signal events can be seen in figure 3.13. This plot is proportional to the efficiency plot, but we can see, that the errors are of considerable magnitude.

Expected Number of Background Events

The background for the $Y(4140)$ signal is expected to consist mainly of $B^{\pm} \to J/\psi\ \phi\ K^{\pm}$ signal events (78.5%, based on the ΔE distribution in figure 3.10) and some background events in the same channel. So, in order to get an estimate for this background, we first performed a fit on the $M(J/\psi\ \phi)$ spectrum in the $B^{\pm} \to J/\psi\ \phi\ K^{\pm}$ signal Monte Carlo, using a phase space inspired function $N \cdot (x - m_1)^a (m_2 - x)^b$, that is intended to describe the $B^{\pm} \to J/\psi\ \phi\ K^{\pm}$ background phenomenologically. The result of this fit can be seen in figure 3.14. We took the obtained shape and fitted it to a generic Monte Carlo sample in order to have the correct normalisation. We finally integrated over the signal area of $4.1\ \text{GeV}/c^2 < M(J/\psi\ \phi) < 4.2\ \text{GeV}/c^2$ on the generic Monte Carlo and thus got the expected number of background events in this area. This whole procedure is repeatedly performed for the various NeuroBayes cuts. The results can be seen in figure 3.15.

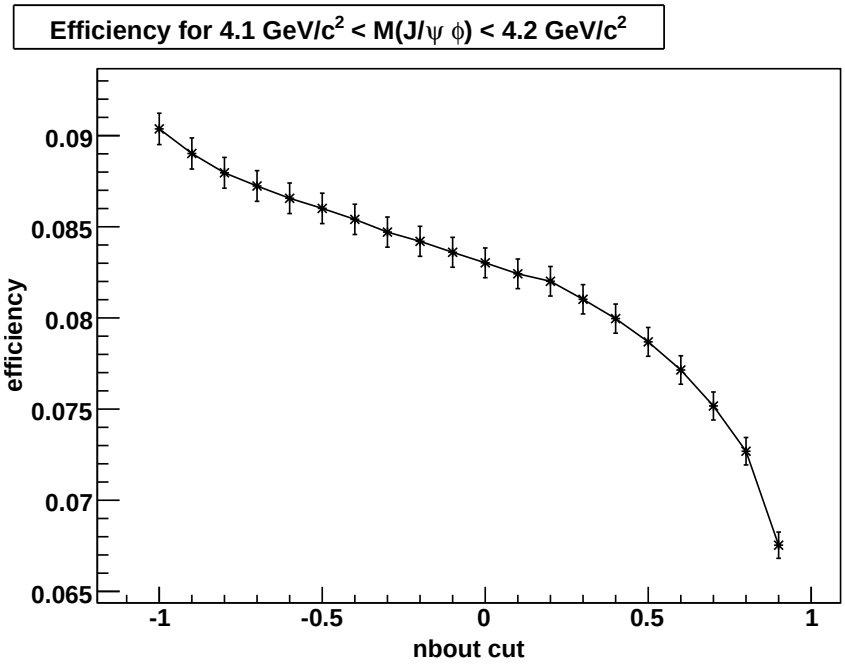


Figure 3.12: The $Y(4140)$ efficiency for events with 4.1 GeV/c^2 $< M(J/\psi\ \phi) <$ 4.2 GeV/c^2.

Expected Significance

With the expected number of signal and background events available, it is now an easy task to calculate the expected significance. We define the significance as

$$\text{significance} = \frac{N(\text{Signal})}{\sqrt{N(\text{Signal}) + N(\text{Background})}} \tag{3.12}$$

Figure 3.16 shows the expected significance depending on the NeuroBayes cut. Although the errors in this plot are rather large due to the uncertainty of the branching ratio of the $Y(4140)$, we can still see that the best significance is to be expected, if we chose a NeuroBayes cut in the region above 0.5. In order to increase the comparability with another, independently developed analysis within the Belle collaboration, we chose a NeuroBayes cut of 0.94. This cut has the advantage to give both analyses similar numbers of signal and background events and delivers an expected significance that is very close to the highest expected significance. With our NeuroBayes cut, we expect 33 signal events, that correspond to a significance of $\sim 3.8\ \sigma$.

3.6.2 Angular Analysis

Depending on the total angular momentum of the $Y(4140)$ it is possible, that the resonance will exist, though we might not be able to detect it, because of the angular distribution of the decay products of the $Y(4140)$. To study this eventuality, the $Y(4140)$

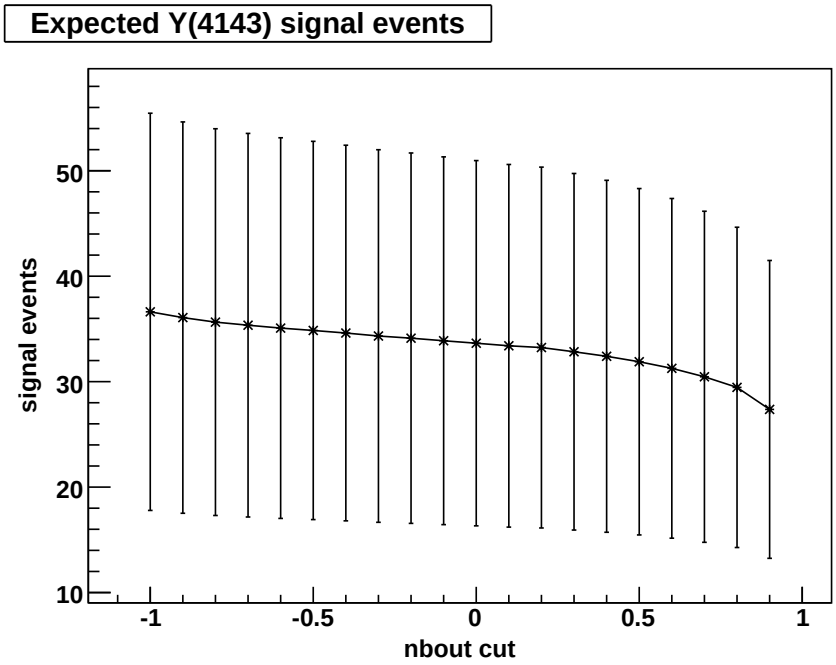


Figure 3.13: The expected number of signal events in the $J/\psi\ \phi$ mass spectrum.

phase space signal Monte Carlo was used again. We studied the efficiency of the reconstruction and selection depending on the helicity angles of J/ψ, ϕ, $K^{\pm}$ from ϕ and leptons from J/ψ. The results can be seen in plots 3.17 - 3.22. It is clearly visible that there are no significant efficiency drops for any angles. So we should be able to see the $Y(4140)$ resonance, if it exists, regardless of the angular distribution of its decay products.

3.7 Analysis on Data

Having the expected significance according to section 3.6.1 and the angular analysis of section 3.6.2 in mind, we expect to see a clear $Y(4140)$ signal peak in the $J/\psi\ \phi$ mass spectrum as soon as we perform the analysis on data. The $B^{\pm} \rightarrow J/\psi\ \phi\ K^{\pm}$ data selection without the $Y(4140)$ resonance is our background. So, in order to obtain a background shape, the selection was performed on the $B^{\pm} \rightarrow J/\psi\ \phi\ K^{\pm}$ signal Monte Carlo and the shape was fitted using a phase space function

$$f(x) = N \cdot (x - m_1)^a \cdot (m_2 - x)^b \ . \tag{3.13}$$

This fit has already been performed to get an estimate for the expected number of background events and the results of this fit can be seen in figure 3.14. The expected signal shape of the $Y(4140)$ signal has already been plotted in figure 3.11, its shape was used for the final fit on data (see figure 3.23). The sum of the fixed signal shape taken

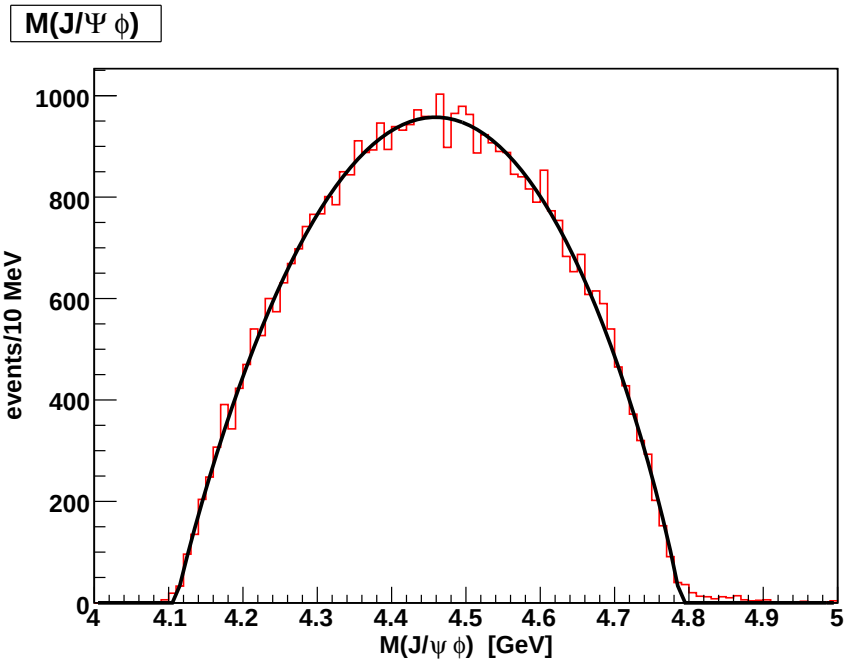


Figure 3.14: The shape of the $M(J/\psi\ \phi)$ spectrum in the $B^{\pm} \to J/\psi\ \phi\ K^{\pm}$ signal Monte Carlo.

from a histogram and the also fixed background shape from the phase space function was then fitted to data using the log-likelihood method. The result of this fit can be seen in figure 3.24. The fit yields 10 ± 5 signal events with 30 ± 2 background events. It has a fit probability of 1.5×10^{-2}. This is a numerical confirmation for what we already see with the naked eye: The $J/\psi\ \phi$ mass spectrum is not well understood and underestimates the background, especially in the region 4.2 GeV/c^2 < $M(J/\psi\ \phi)$ < 4.4 GeV/c^2. Our assumption of a peaking structure at $\sim$ 4140 MeV/c^2 accidentally barely fits such an elevated background level. Another, clearer "peak" can be found at a slightly higher energy and it is completely neglected in the fit in figure 3.24. We therefore made another fit, where the only fixed component of the phase-space background was the threshold mass $m_1 = 4.1$ GeV/c^2 (see equation 3.13). The rest of the background shape was left free. This seems reasonable, as the background is not well-described by the Monte Carlo samples. The result of the new fit can bee seen in figure 3.25. It has a fit probability of 9.1×10^{-2}, which is a factor of $\sim$ 6 higher than in the previous fit, and gives back 6 ± 6 signal events and 33 ± 16 background events. These results are well compatible with no signal and the fit has a far better probability than the fit with the fixed but not understood background shape. We therefore deem the $Y(4140)$ resonance to be not observed within this analysis.

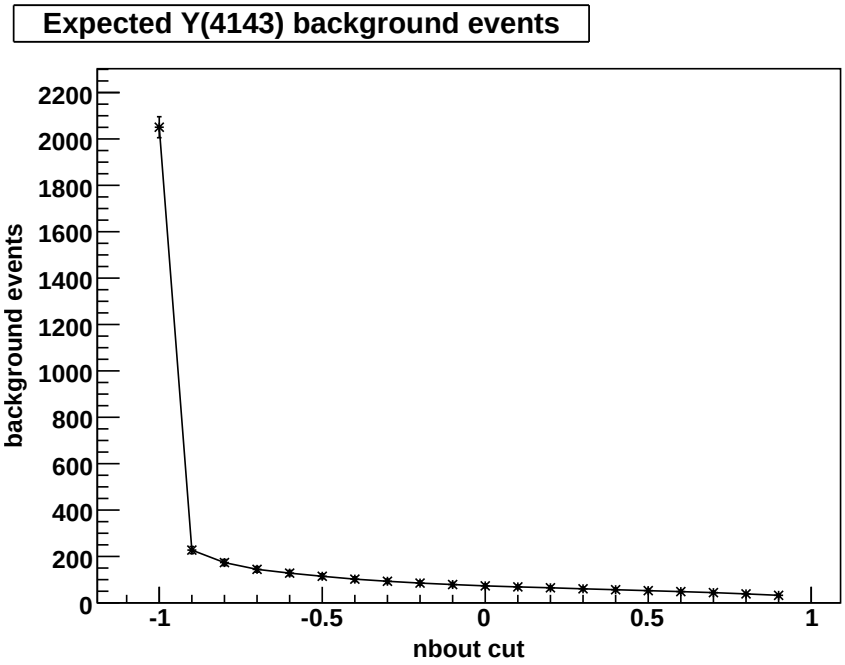


Figure 3.15: The expected number of background events in the $Y(4140)$ signal region.

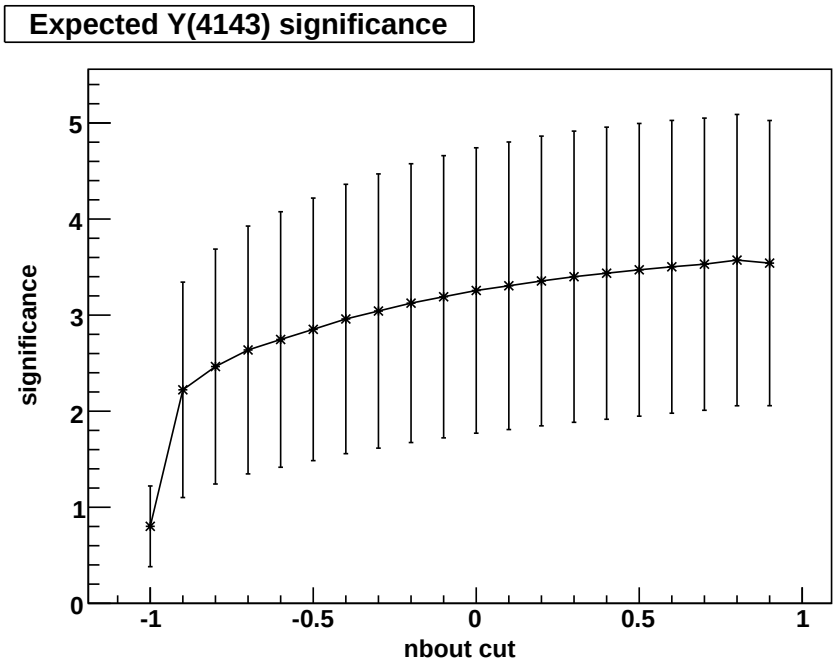


Figure 3.16: The expected significance of the $Y(4140)$ signal.

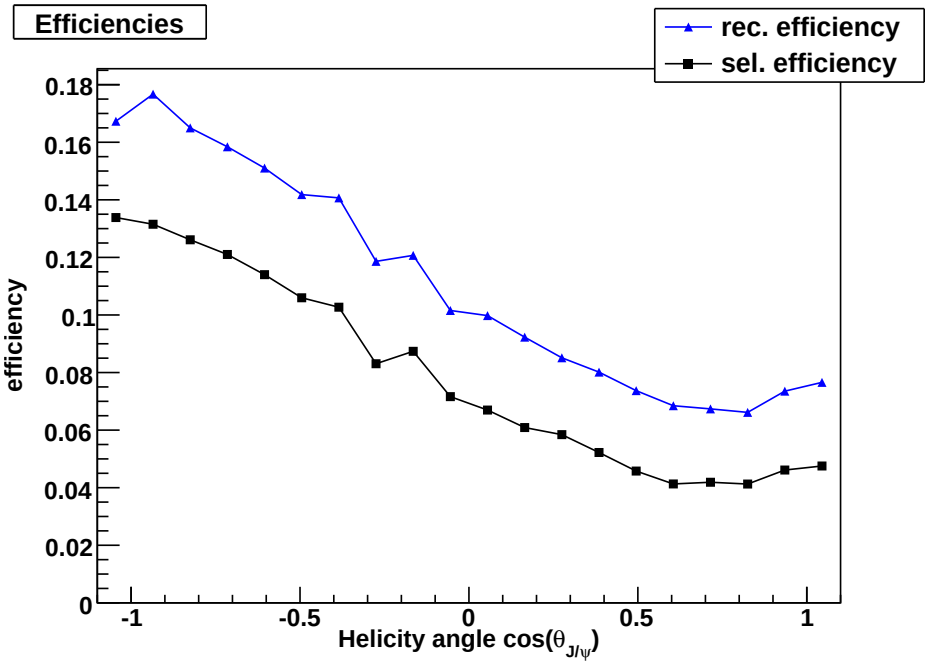


Figure 3.17: Efficiency vs helicity angle of the J/ψ.

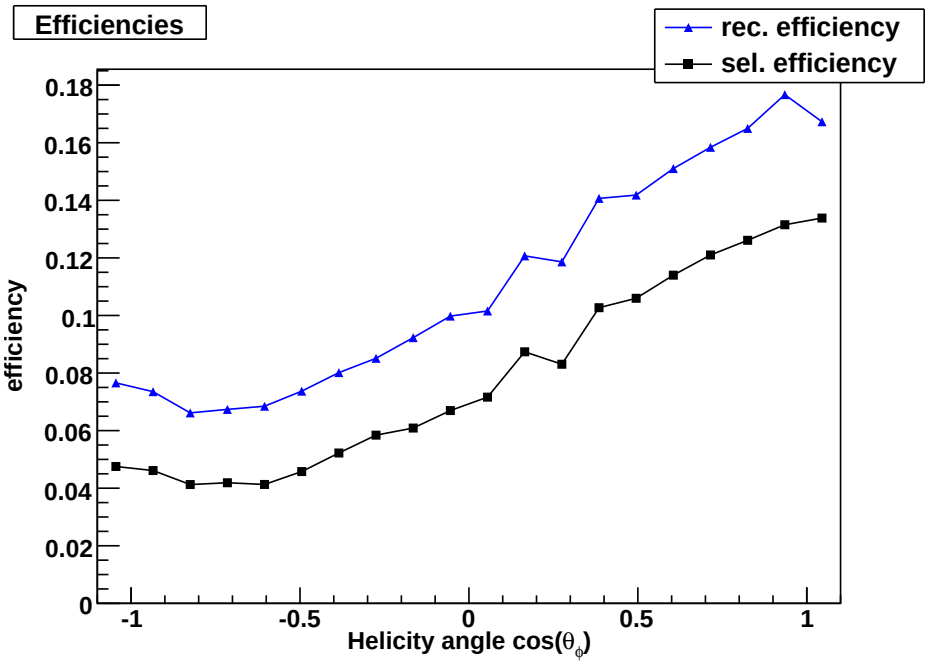


Figure 3.18: Efficiency vs helicity angle of the ϕ.

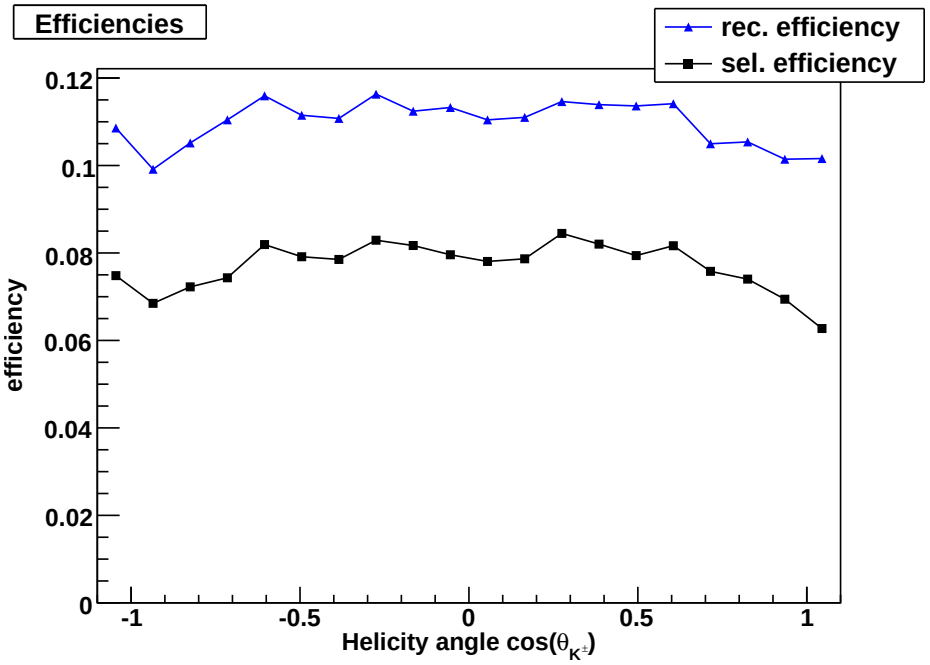


Figure 3.19: Efficiency vs helicity angle of the $K^{\pm}$.

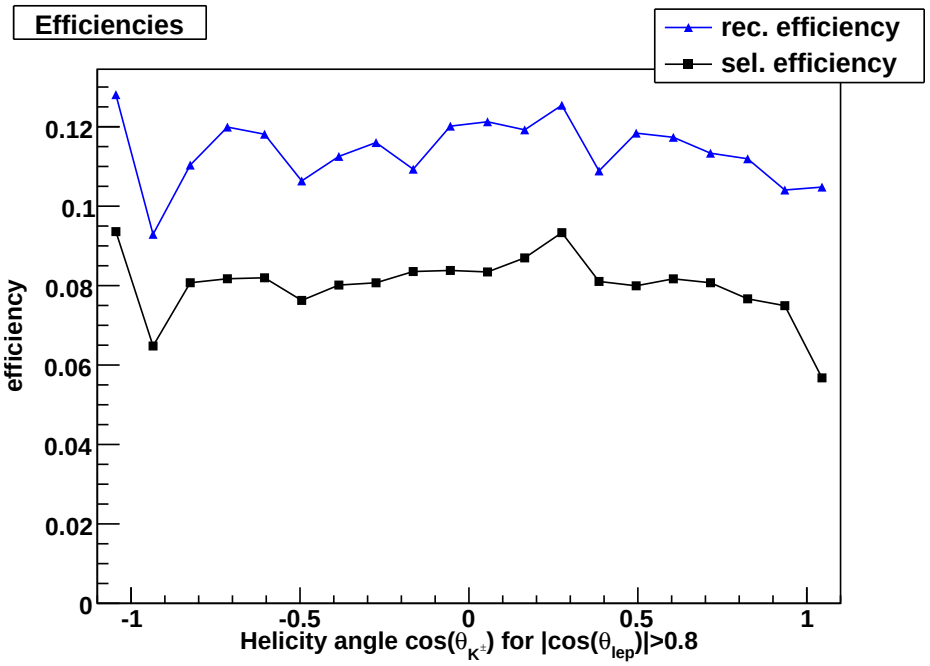


Figure 3.20: Efficiency vs helicity angle of the $K^{\pm}$ for large lepton angles.

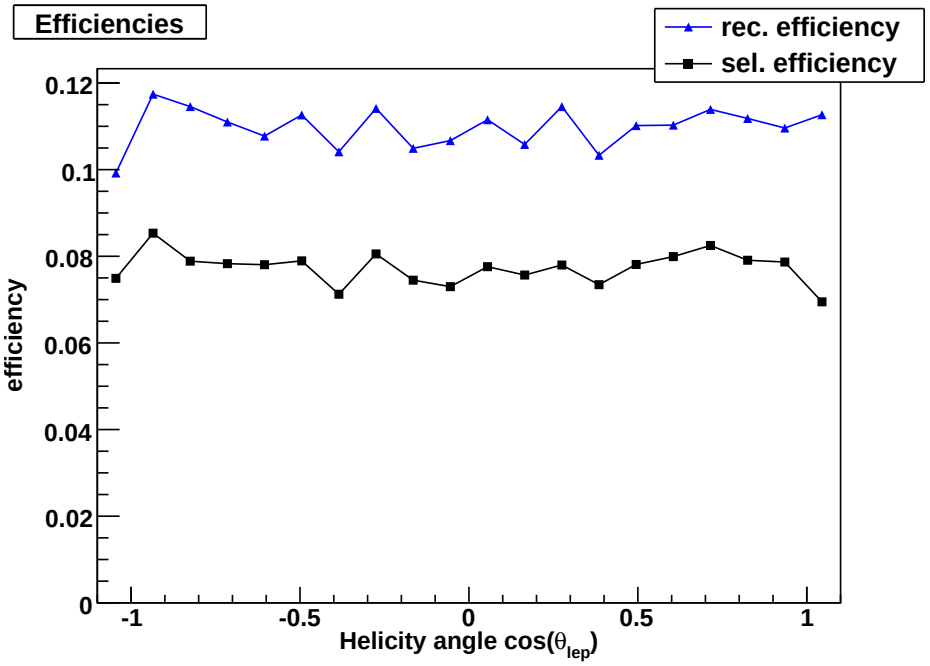


Figure 3.21: Efficiency vs helicity angle of the leptons from J/ψ.

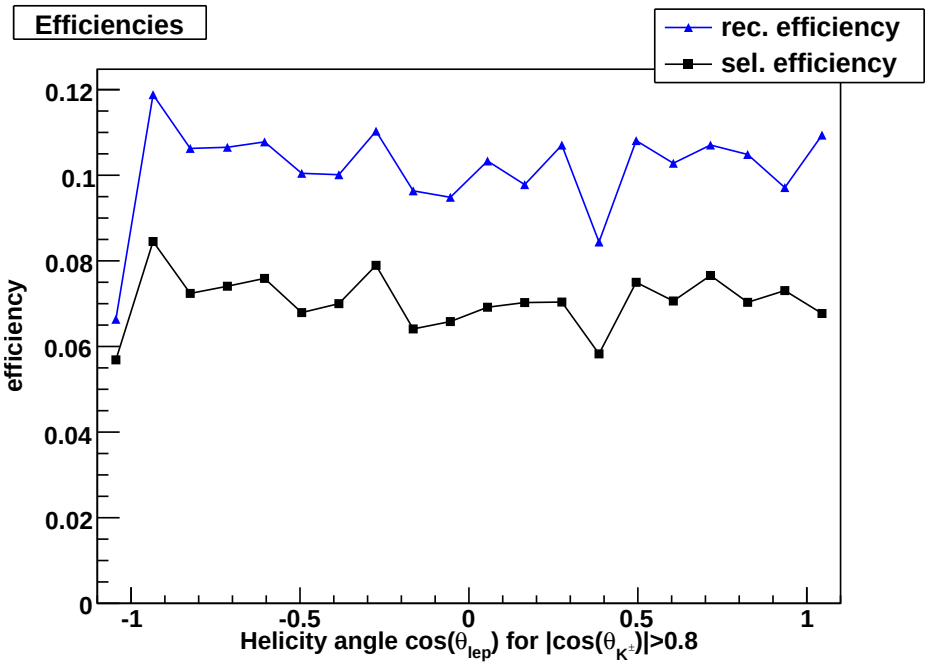


Figure 3.22: Efficiency vs helicity angle of the leptons from J/ψ for large $K^{\pm}$ angles.

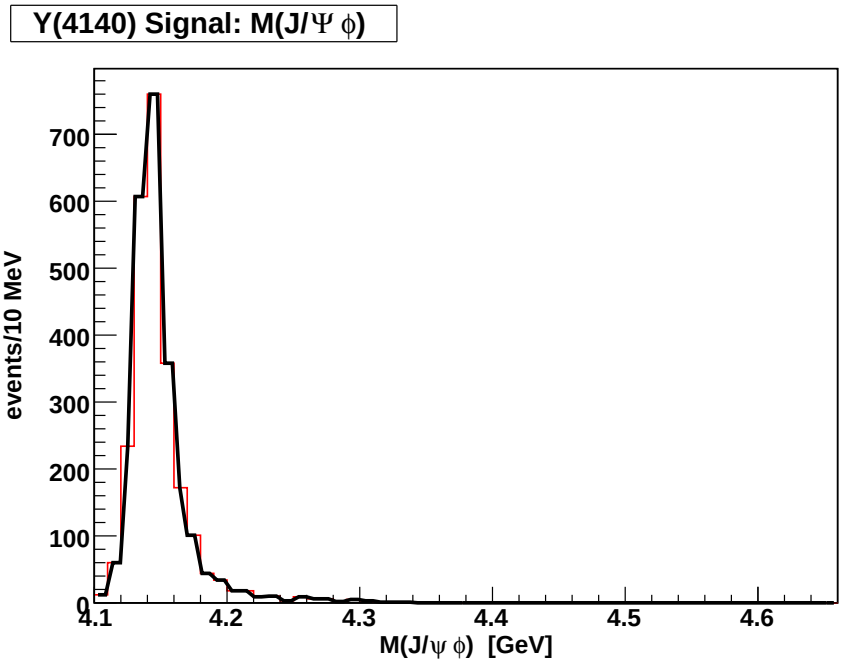


Figure 3.23: Expected shape of the $Y(4140)$ signal.

3.7.1 Upper Limit

To quantify our result, we calculated an Bayesian upper limit with a flat prior of the branching ratio of the $Y(4140)$ resonance. To do so, the number of signal events in the final maximum likelihood fit was fixed at different levels. During this procedure, the background shape was allowed to be adapted within the fit, except the for the threshold value that was fixed to $m_1 = 4.1$ GeV/c^2 as on the fit on data (see equation 3.13). For each one of these different fits, the fit probability was plotted over the given number of signal events. The result can be seen in figure 3.26. An integration over histogram 3.26 up until a certain percentage x of the whole area under the histogram gives an upper limit on the number of signal events with the before used percentage x as confidence level. We chose to integrate up until 90% and calculated the branching ratio, that corresponds to the obtained 18 signal entries. This was done using equation 3.11:

$$\mathrm{BR}(Y(4140)) = \frac{N_{Y(4140)}}{\underbrace{771 \times 10^6}_{\#B\bar{B}} \cdot \underbrace{0.0619}_{\text{efficiency}} \cdot \underbrace{0.1187 \cdot 0.492}_{J/\psi \text{ and } \phi \text{ BR}}} \ . \tag{3.14}$$

We obtained a branching ratio for the $Y(4140)$ resonance of

$$\mathrm{BR}(Y(4140)) < 6.6 \times 10^{-6} \ , \tag{3.15}$$

with 90% confidence level.

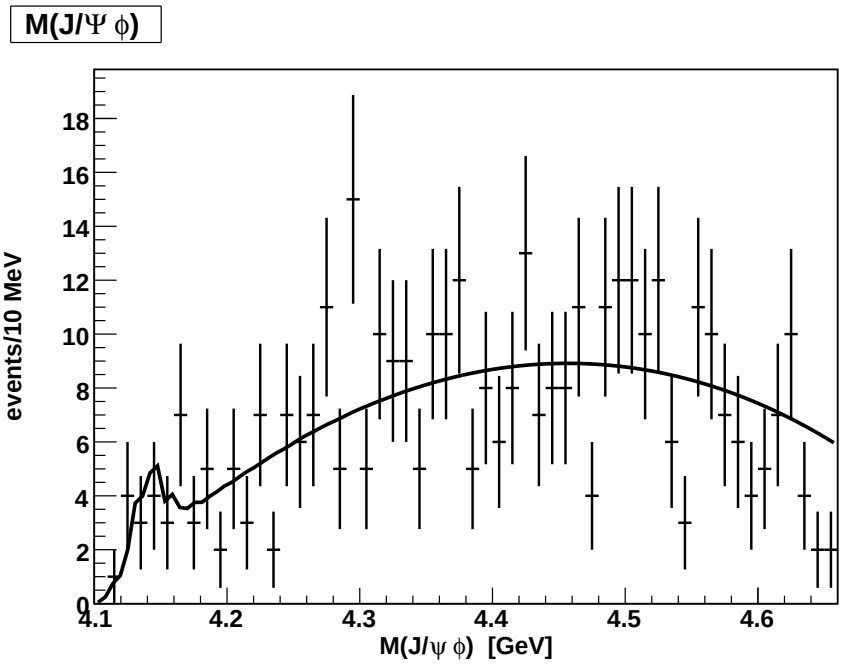


Figure 3.24: $J/\psi\phi$ mass spectrum on the $B^{\pm} \rightarrow J/\psi\ \phi\ K^{\pm}$ selection on data, fixed background shape.

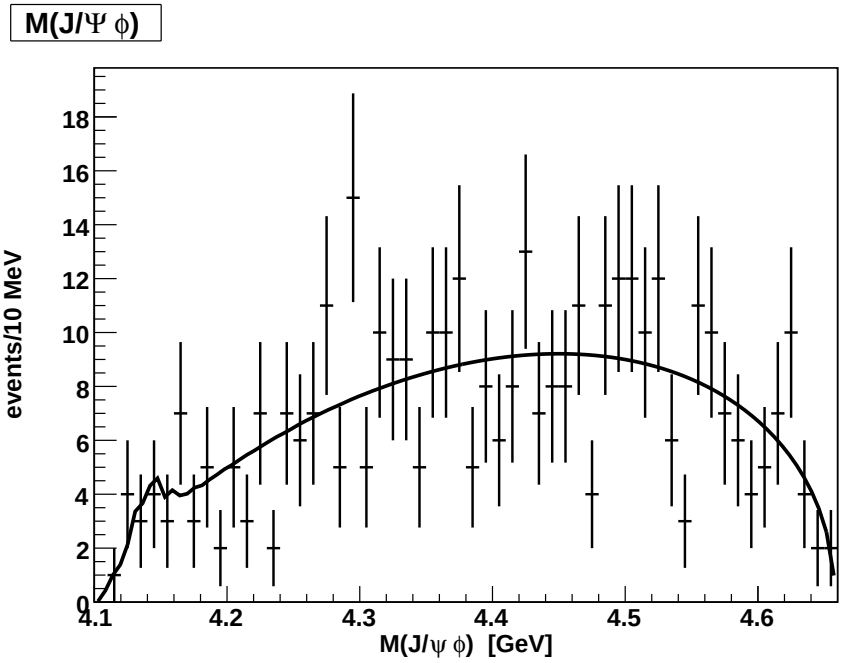


Figure 3.25: $J/\psi\phi$ mass spectrum on the $B^{\pm} \rightarrow J/\psi\ \phi\ K^{\pm}$ selection on data, open background shape.

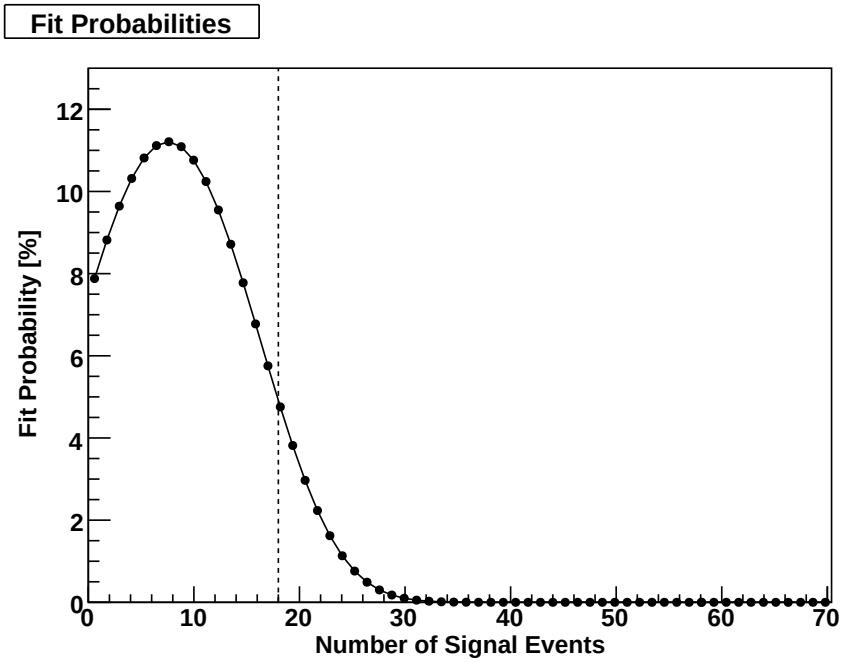


Figure 3.26: Fit Probabilities for different numbers of signal events.

4 Conclusion

The presented thesis describes the first attempts to introduce the usage of neural network techniques in Belle analyses.

Full Reconstruction

The most involved project will be the full reconstruction technique, whose foundations were built as part of this thesis, and whose first results were also presented. Utilising the multivariate analysis tool NeuroBayes, the identification of charged tracks could be improved. The main challenge will be the reconstruction and classification of hundreds of different B decay channels, utilising multiple hundreds of NeuroBayes experts in reasonable programming and execution time.

$Y(4140)$

A newsworthy analysis of the CDF collaboration was then tested on Belle data. The CDF collaboration reported to have found evidence for a narrow structure in the $J/\psi\ \phi$ mass spectrum at a mass of ~ 4143 MeV. This so called $Y(4140)$ resonance was considered a candidate for a tetraquark. The analysis described in this thesis was done utilising the architecture of the full reconstruction program and thereby testing the design for major flaws. We reconstructed the decay channel

$$B^{\pm} \to J/\psi\ \phi\ K^{\pm} \tag{4.1}$$

in search for the $Y(4140)$ resonance in the $J/\psi\ \phi$ mass spectrum at an energy of ~ 4140 MeV. After the reconstruction process, an analysis was performed, during which the NeuroBayes package was used predominantly for statistical analysis. After the selection process, the final fit on data yielded 6 ± 6 signal events. Assuming the properties of the resonance to be correctly measure by the CDF collaboration, we expected ~ 30 signal events. We thus came to the conclusion that the $Y(4140)$ resonance could not be found in Belle data. The upper limit on the branching ratio was calculated to be

$$\mathrm{BR}(Y(4140)) < 6.6 \times 10^{-6}\ .$$

In addition to the analysis described in this thesis, another $Y(4140)$ analysis was done by members of the Belle collaboration on the same topic in parallel, but with a different analysis technique. This provided valuable information for comparison of the different techniques. The other Belle analysis obtained consistent and comparable results and was also not able to find a significant signal.

A List of Variables in the Neural Networks

Below follow tables with the variables that were used for the training of the respective neural networks. Whenever a variable appears multiple times in such a table, it is due to different preprocessings.

A.1 $e^{\pm}$ Neural Network

No.	Variable	add sig.	only this	loss	correl
1	Electron ID	206.87	206.87	89.98	84.4
2	Electron ID	85.59	163.72	51.51	69.0
3	`TOF PID` for $\pi^{\pm}$	31.56	155.69	22.70	69.7
4	Electron ID	16.20	157.96	17.88	80.5
5	`TOF PID` for electrons	18.16	148.11	18.18	68.1
6	Number of photo electrons	11.44	69.32	10.38	58.4
7	`TOF PID` for $p^{\pm}$	12.45	9.53	13.87	65.1
8	Number of hits in the stereo wire	10.78	34.13	4.05	69.0
9	TOF Conf. level for electron hypothesis	10.02	1.24	9.06	70.4
10	`ACC PID` for $p^{\pm}$	10.16	56.95	6.35	81.8
11	p_x	6.86	10.49	5.42	55.5
12	dE/dx for $p^{\pm}$ hypothesis	6.64	57.47	3.63	98.5
13	`TOF PID` for $\pi^{\pm}$	6.03	9.73	5.32	47.9
14	p_z	4.53	23.59	4.41	53.0
15	`CDC PID` for $K^{\pm}$	3.94	62.00	3.85	49.8
16	p_y	2.11	33.54	2.02	55.8
17	dE/dx for $e^{\pm}$ hypothesis	1.96	57.86	1.98	98.5
18	`ACC PID` for $p^{\pm}$	1.26	49.25	1.26	81.2
19	Path length from TOF system	0.22	8.22	0.22	33.2
20	TOF Conf. level for electron hypothesis	0.04	28.99	0.04	62.6

Table A.1: Variable list for the $e^{\pm}$ neural network.

A.2 $\mu^{\pm}$ Neural Network

No.	Variable	add sig.	only this	loss	correl
1	Flag, if std. muon ID exists	53.65	53.65	51.18	31.9
2	Std. muon ID	43.07	43.10	25.97	63.5
3	dE/dx for $\pi^{\pm}$ hypothesis	35.86	35.15	11.29	73.2
4	Starting y of the track near the 1st CDC hit	29.52	31.84	16.02	60.2
5	`ATC PID` for $\mu^{\pm}$ vs. $\pi^{\pm}$	21.78	35.75	22.57	40.6
6	Starting x of the track near the 1st CDC hit	13.35	29.81	13.35	60.2
7	No. of hits in KLM	9.53	37.35	10.05	64.1
8	dE/dx for $\mu^{\pm}$ hypothesis	4.95	26.52	10.82	82.5
9	dE/dx	8.00	13.38	9.18	80.8
10	dE/dx for $p^{\pm}$ hypothesis	8.56	9.40	10.64	50.2
11	`TOF PID` for $\pi^{\pm}$	9.16	19.65	6.93	57.8
12	`ATC PID` for $\pi^{\pm}$ vs. $e^{\pm}$	7.17	19.21	6.64	99.7
13	Flag, if std. electron ID exists	6.34	7.14	6.69	48.4
14	`TOF PID` for $K^{\pm}$	6.07	7.42	6.12	20.9
15	`ATC PID` for $e^{\pm}$ vs. $\pi^{\pm}$	6.05	19.13	6.05	99.7

Table A.2: Variable list for the $\mu^{\pm}$ neural network.

A.3 $\pi^{\pm}$ Neural Network

No.	Variable	add sig.	only this	loss	correl
1	`ATC PID` for $\pi^{\pm}$ vs. $K^{\pm}$	89.91	89.91	3.24	99.6
2	Electron ID	56.72	64.38	20.68	63.9
3	`ATC PID` for $\mu^{\pm}$ vs. $\pi^{\pm}$	31.84	57.35	17.10	61.4
4	Muon ID	24.17	22.31	22.55	14.1
5	Flag, if std. muon ID exists	21.73	22.23	24.06	29.4
6	`TOF PID` for $\pi^{\pm}$	20.03	81.95	22.22	89.8
7	`TOF PID` for $\mu^{\pm}$	10.79	51.42	14.71	87.6
8	dE/dx for $e^{\pm}$ hypothesis	13.49	47.62	13.64	58.4
9	Starting y of the track near the 1st CDC hit	11.21	17.48	6.52	61.2
10	TOF Conf. level for $K^{\pm}$ hypothesis	11.05	47.07	9.66	47.6
11	`ATC PID` for $K^{\pm}$ vs. $p^{\pm}$	9.63	56.37	11.22	64.0
12	`ATC PID` for $K^{\pm}$ vs. $e^{\pm}$	8.96	61.66	9.63	72.3
13	Starting x of the track near the 1st CDC hit	5.81	16.83	6.00	61.3
14	dE/dx for $p^{\pm}$ hypothesis	5.04	59.50	4.98	69.8
15	`ATC PID` for $\pi^{\pm}$ vs. $K^{\pm}$	1.64	89.54	1.64	99.6

Table A.3: Variable list for the $\pi^{\pm}$ neural network.

A.4 $K^{\pm}$ Neural Network

No.	Variable	add sig.	only this	loss	correl
1	`ATC PID` for $\pi^{\pm}$ vs. $K^{\pm}$	105.46	105.46	16.66	99.0
2	`ATC PID` for $e^{\pm}$ vs. $K^{\pm}$	18.21	92.57	8.72	84.0
3	`TOF PID` for $e^{\pm}$	12.57	75.68	17.39	79.8
4	TOF Conf. level for $K^{\pm}$ hypothesis	8.92	59.00	9.11	58.4
5	dE/dx for $p^{\pm}$ hypothesis	8.59	32.50	4.49	93.7
6	`ACC PID` for $p^{\pm}$	7.06	18.90	7.54	48.1
7	`ATC PID` for $\pi^{\pm}$ vs. $p^{\pm}$	4.09	58.87	5.74	79.8
8	`ATC PID` for $\pi^{\pm}$ vs. $K^{\pm}$	5.43	103.63	6.24	99.1
9	`TOF PID` for $\mu^{\pm}$	4.22	33.87	4.56	54.2
10	`ATC PID` for $p^{\pm}$ vs. $e^{\pm}$	3.24	44.21	3.22	72.4
11	Flag, if std. muon ID exists	3.15	3.68	4.34	89.8
12	`KLM PID` for $\mu^{\pm}$	1.67	6.29	3.38	65.4
13	`KLM PID` for $\pi^{\pm}$	3.21	14.42	3.27	90.5
14	Starting z of the track near the 1st CDC hit	2.94	13.45	2.69	34.4
15	dE/dx for $K^{\pm}$ hypothesis	2.12	48.00	2.12	95.5

Table A.4: Variable list for the $K^{\pm}$ neural network.

A.5 J/ψ Neural Network

No.	Variable	add sig.	only this	loss	correl
1	NB output of the second child	358.98	358.98	7.97	99.8
2	χ^2 of the vertex fit	180.02	324.65	45.83	90.2
3	Mass before mass contrained veretx fit	56.08	314.89	53.00	80.3
4	Product of children NB outputs (transformed to $[0, 1]$)	11.89	240.86	18.15	92.3
5	Electron-muon flag	18.64	10.08	20.82	77.0
6	p_z before vertex fit	18.54	64.40	10.19	70.6
7	CL of the vertex fit	20.33	275.14	19.15	81.8
8	Sum of children NB outputs	10.94	235.81	7.41	91.8
9	Invariant mass of added radiated photons	10.65	24.54	10.31	17.2
10	Mass after mass constrained vertex fit	10.38	191.36	8.15	66.1
11	Cosine of the angle between the children in the lab frame	7.66	41.70	5.95	42.2
12	p_x	7.74	39.10	6.59	37.2
13	Product of children NB outputs (transformed to $[0, 1]$)	6.75	358.62	6.85	99.8
14	Energy	4.81	61.62	4.03	77.4
15	p_z of second child before vertex fit	3.37	47.26	4.34	62.7
16	Energy of first child before vertex fit	2.48	6.44	4.17	94.0
17	Energy of second child before vertex fit	2.70	8.76	3.54	93.8
18	Energy before vertex fit	2.04	9.19	2.36	84.0
19	NB output of the first child	1.27	127.60	1.31	96.1
20	NB output of the second child	2.00	115.37	2.01	94.7
21	NB output of the first child	1.30	115.75	1.29	94.8
22	Mass before mass constrained vertex fit	0.78	57.89	0.78	59.6

Table A.5: Variable list for the J/ψ neural network.

A.6 ϕ Neural Network

No.	Variable	add sig.	only this	loss	correl
1	Sum of children NB outputs	373.86	373.86	18.05	99.0
2	Mass pull	138.19	273.45	0.00	99.6
3	Mass before vertex fit	50.47	276.82	38.56	85.9
4	Energy	29.20	147.51	0.00	99.9
5	Product of children NB outputs (transformed to $[0,1]$)	20.69	375.18	21.01	98.9
6	Mass	16.74	278.55	13.21	94.8
7	$\|p\|$	16.39	152.97	10.85	99.9
8	Cosine of the angle between the children in the lab frame	14.04	161.60	13.27	89.2
9	p_x of the first child	4.37	116.44	3.16	98.1
10	p_x of the second child	1.52	115.07	1.74	98.1
11	p_x before vertex fit	7.86	128.75	0.00	99.8
12	p_y of the first child	6.85	118.37	1.56	98.3
13	p_y of the second child	3.84	117.52	0.00	98.3
14	p_y before vertex fit	9.17	131.14	0.00	99.8
15	ΔE before vertex fit	9.30	149.01	0.00	99.9
16	p_z of the second child	6.92	87.10	3.35	98.0
17	p_z of the first child	6.17	87.46	2.36	98.2
18	P_z before vertex fit	9.36	94.17	6.21	99.7
19	χ^2 of vertex fit	7.93	98.76	0.00	78.9
20	p_z	6.25	95.01	0.00	99.6
21	NB output of the first child	5.40	184.93	0.00	70.1
22	NB output of the second child	5.63	182.82	0.00	70.3
23	Mass error	4.40	69.61	0.00	69.6
24	Energy of the second child	2.51	138.15	0.00	100.0
25	Energy of the first child	3.26	139.38	0.00	100.0
26	p_x	3.75	129.09	0.00	99.8
27	p_y	3.53	131.50	0.00	99.8
28	Cosine of the angle between the children in the CMS of the ϕ	3.43	14.15	0.00	68.4
29	Energy before vertex fit	1.05	147.21	0.00	99.9
30	CL of vertex fit	0.61	75.64	0.00	77.4
31	Absolute value of the mass pull	0.55	273.44	0.00	99.6
32	$\|p\|$ of the second child	0.00	138.15	0.00	100.0
33	$\|p\|$ of the first child	0.00	139.38	0.00	100.0
34	M_{BC} before the vertex fit	0.00	154.26	0.00	99.9

Table A.6: Variable list for the ϕ neural network.

A.7 $B^{\pm}$ Neural Network

No.	Variable	add sig.	only this	loss	correl
1	Product of children NB outputs (transformed to $[0,1]$)	9.70	489.70	66.77	95.6
2	M_{BC} after mass constrained vertex fits for J/ψ and B	8.38	470.08	52.42	92.5
3	M_{BC} before any mass constrained vertex fits	8.07	467.27	31.36	95.3
4	Sum of children NB outputs	8.07	475.94	16.79	96.3
5	M_{BC} before any mass constrained vertex fits	7.81	290.59	6.45	93.7
6	NB output of the ϕ	4.56	355.23	13.21	84.4
7	M_{BC}	2.32	465.35	11.89	96.1
8	Energy of the ϕ before any vertex fits	1.65	21.13	10.89	71.2
9	NB output for the J/ψ	1.49	243.56	5.42	85.6
10	Energy of the $K^{\pm}$	0.10	93.93	10.40	24.3
11	Energy of the ϕ before any vertex fits	7.14	111.11	7.04	78.2
12	Energy of the ϕ after mass constrained vertex fits for J/ψ and B	7.94	85.98	6.54	76.7
13	NB output of the ϕ	5.58	258.15	5.23	74.6
14	NB output of the $K^{\pm}$	5.13	234.93	3.47	78.9
15	p_x of the ϕ before any vertex fits	4.53	72.61	4.83	39.2
16	p_y of the ϕ before any vertex fits	4.91	72.73	4.60	40.3
17	Mass of the ϕ before any vertex fits	3.14	120.44	3.37	52.1
18	M_{BC} after mass constrained vertex fits for J/ψ and B	3.09	293.84	2.77	94.6
19	p_z of the ϕ before any vertex fits	3.13	0.68	3.00	68.0
20	p_y of the J/ψ before any vertex fits	2.94	64.20	3.02	37.5
21	Mass of the J/ψ before any vertex fits	2.73	140.99	2.01	60.7
22	p_x of the J/ψ after mass constrained vertex fits for J/ψ and B	2.07	52.51	2.08	33.8
23	NB output for the J/ψ	2.07	197.16	1.85	82.7
24	Cosine of the angle between J/ψ and ϕ in the $B^{\pm}$ CMS	1.60	6.69	1.63	54.2
25	Cosine of the angle between J/ψ and ϕ in the $B^{\pm}$	1.55	44.39	1.50	37.5
26	χ^2 of vertex fit	1.15	175.30	1.11	40.6
27	Mass of the J/ψ after mass constrained vertex fits for J/ψ and B	0.98	117.58	0.98	55.5
28	Cosine of the angle between J/ψ and $K^{\pm}$ in the lab frame	0.89	51.46	0.92	22.6

29	NB output of the $K^{\pm}$	0.39	299.73	0.39	84.7
30	M_{BC}	0.15	291.99	0.15	96.1

Table A.7: Variable list for the $B^{\pm}$ neural network.

B Channels for the Full Reconstruction

B.1 B^0 Modes

Channel	Γ_i/Γ	
$B^0 \to D^-\ \pi^+$	(3.4 ± 0.9)	$\times 10^{-3}$
$B^0 \to D^-\ K^+$	(2.0 ± 0.6)	$\times 10^{-4}$
$B^0 \to D^-\ \rho^+$	(7.5 ± 1.2)	$\times 10^{-3}$
$B^0 \to D^-\ a_1(1260)^+$	(6.0 ± 3.3)	$\times 10^{-4}$
$B^0 \to D^*(2010)^-\ \pi^+$	(2.76 ± 0.21)	$\times 10^{-3}$
$B^0 \to D^*(2010)^-\ K^+$	(2.14 ± 0.20)	$\times 10^{-4}$
$B^0 \to D^*(2010)^-\ \rho^+$	(6.8 ± 0.9)	$\times 10^{-3}$
$B^0 \to D^*(2010)^-\ a_1(1260)^+$	(1.30 ± 0.27)	$\times 10^{-2}$
$B^0 \to D^-\ D_S^+$	(6.5 ± 2.1)	$\times 10^{-3}$
$B^0 \to D^-\ D_S^{*+}$	(8.6 ± 3.4)	$\times 10^{-3}$
$B^0 \to D^*(2010)^-\ D_S^+$	(8.8 ± 1.6)	$\times 10^{-3}$
$B^0 \to D^*(2010)^-\ D_S^{*+}$	(1.79 ± 0.16)	$\times 10^{-2}$
$B^0 \to J/\psi(1S)\ K^0$	(8.27 ± 0.33)	$\times 10^{-4}$
$B^0 \to J/\psi(1S)\ K^*(892)^0$	(1.33 ± 0.06)	$\times 10^{-3}$
$B^0 \to \psi(2S)\ K^0$	(6.2 ± 0.6)	$\times 10^{-4}$
$B^0 \to \psi(2S)\ K^*(892)^0$	(7.2 ± 0.8)	$\times 10^{-4}$
$B^0 \to \eta\ K^*(892)^0$	(1.77 ± 0.23)	$\times 10^{-5}$
$B^0 \to \eta'\ K^0$	(6.8 ± 0.4)	$\times 10^{-5}$
$B^0 \to K^*(892)^0\ \gamma$	(4.01 ± 0.20)	$\times 10^{-5}$
$B^0 \to K^*(892)^+\ \pi^-$	(1.18 ± 0.15)	$\times 10^{-5}$
$B^0 \to K^+\ \pi^-$	(1.82 ± 0.08)	$\times 10^{-5}$
$B^0 \to \bar{D}^0\ \pi^0$	(2.91 ± 0.28)	$\times 10^{-4}$
$B^0 \to \bar{D}^0\ \eta$	(2.2 ± 0.5)	$\times 10^{-4}$
$B^0 \to \bar{D}^0\ \eta'$	(1.25 ± 0.23)	$\times 10^{-4}$
$B^0 \to D^-\ K^*(892)^+$	(4.5 ± 0.7)	$\times 10^{-4}$
$B^0 \to D^*(2010)^-\ K^*(892)^+$	(3.3 ± 0.6)	$\times 10^{-4}$
$B^0 \to D^-\ D^+$	(1.9 ± 0.6)	$\times 10^{-4}$
$B^0 \to D^-\ K^+\ \bar{K}*(892)^0$	(8.8 ± 1.9)	$\times 10^{-4}$
$B^0 \to \bar{D}^0\ \pi^+\ \pi^-$	(8.0 ± 1.6)	$\times 10^{-4}$
$B^0 \to D^*(2010)^-\ \pi^+\ \pi^0$	(1.5 ± 0.5)	$\times 10^{-2}$
$B^0 \to D^-\ D^*(2007)^0\ K^+$	(4.6 ± 1.0)	$\times 10^{-3}$
$B^0 \to D^*(2010)^-\ D^0\ K^+$	$(3.1^{+0.6}_{-0.5})$	$\times 10^{-3}$
$B^0 \to D^*(2010)^-\ D^*(2007)^0\ K^+$	(1.18 ± 0.20)	$\times 10^{-2}$

Channel	Γ_i/Γ	
$B^0 \to D^*(2010)^-\ D^*(2010)^+\ K^0$	(8.8 ± 1.9)	$\times 10^{-3}$
$B^0 \to D^-\ \omega\ \pi^+$	(2.8 ± 0.6)	$\times 10^{-3}$
$B^0 \to D^*(2010)^-\ \omega\ \pi^+$	(2.9 ± 0.5)	$\times 10^{-3}$
$B^0 \to D^-\ \pi^+\ \pi^+\ \pi^-$	(8.0 ± 2.5)	$\times 10^{-3}$
$B^0 \to D^*(2010)^-\pi^+\pi^+\pi^-$	(7.0 ± 0.8)	$\times 10^{-3}$
$B^0 \to D^*(2010)^-\ \pi^+\ \pi^+\ \pi^-\ \pi^0$	(1.76 ± 0.27)	$\times 10^{-2}$
$B^0 \to D^{*-}\ 3\pi^+\ 2\pi^-$	(4.7 ± 0.9)	$\times 10^{-3}$

B.2 B^- Modes

Channel	Γ_i/Γ	
$B^- \to D^0\ \pi^-$	(4.92 ± 0.20)	$\times 10^{-3}$
$B^- \to D^0\ K^-$	(4.08 ± 0.24)	$\times 10^{-4}$
$B^- \to D^0\ \rho^-$	(1.34 ± 0.18)	$\times 10^{-2}$
$B^- \to D^0\ a_1(1260)^-$	(4 ± 4)	$\times 10^{-3}$
$B^- \to D^*(2007)^0\ \pi^-$	(4.6 ± 0.4)	$\times 10^{-3}$
$B^- \to D^*(2007)^0\ K^-$	(3.7 ± 0.4)	$\times 10^{-4}$
$B^- \to D^*(2007)^0\ \rho^-$	(9.8 ± 1.7)	$\times 10^{-3}$
$B^- \to D^*(2007)^0\ a_1(1260)^-$	(1.9 ± 0.5)	$\times 10^{-2}$
$B^- \to D^0\ D_S^+$	(1.09 ± 0.27)	$\times 10^{-2}$
$B^- \to D^0\ D_S^{*+}$	(7.2 ± 2.6)	$\times 10^{-3}$
$B^- \to D^*(2007)^0\ D_S^-$	(10 ± 4)	$\times 10^{-3}$
$B^- \to D^*(2007)^0\ D_S^{*-}$	(2.2 ± 0.7)	$\times 10^{-2}$
$B^- \to J/\psi(1S)\ K^-$	(1.008 ± 0.035)	$\times 10^{-3}$
$B^- \to J/\psi(1S)\ K^*(892)^-$	(1.41 ± 0.08)	$\times 10^{-3}$
$B^- \to \psi(2S)\ K^-$	(6.48 ± 0.35)	$\times 10^{-4}$
$B^- \to \psi(2S)\ K^*(892)^-$	(6.7 ± 1.4)	$\times 10^{-4}$
$B^- \to \eta\ K^*(892)^-$	(2.6 ± 0.4)	$\times 10^{-5}$
$B^- \to \eta'\ K^-$	(7.05 ± 0.35)	$\times 10^{-5}$
$B^- \to K^*(892)^-\ \gamma$	(4.03 ± 0.26)	$\times 10^{-5}$
$B^- \to K^*(892)^0\ \pi^-$	(1.16 ± 0.16)	$\times 10^{-5}$
$B^- \to K^0\ \pi^-$	(2.41 ± 0.17)	$\times 10^{-5}$
$B^- \to D^0\ K^*(892)^-$	(6.3 ± 0.8)	$\times 10^{-4}$
$B^- \to D^*(2007)^0\ K^*(892)^-$	(8.1 ± 1.4)	$\times 10^{-4}$
$B^- \to D^0\ D^-$	(4.8 ± 1.0)	$\times 10^{-4}$
$B^- \to D^0_{CP}\ D^-$	$(\pm)$	$\times 10^{-}$
$B^- \to D^0\ D^*(2910)^-$	(4.6 ± 0.9)	$\times 10^{-4}$
$B^- \to D^0\ K^-\ K^0$	(5.5 ± 1.6)	$\times 10^{-4}$
$B^- \to D^0\ K^-\ K^*(892)^0$	(7.5 ± 1.7)	$\times 10^{-4}$

Channel	Γ_i/Γ	
$B^- \to D^*(2007)^0\ K^-\ K^*(892)^0$	(1.5 ± 0.4)	$\times 10^{-3}$
$B^- \to D^+\ \pi^-\ \pi^-$	(1.02 ± 0.16)	$\times 10^{-3}$
$B^- \to D^*(2010)^+\ \pi^-\ \pi^-$	(1.35 ± 0.22)	$\times 10^{-3}$
$B^- \to D^0\ D^*(2010)^-\ K^0$	(5.2 ± 1.2)	$\times 10^{-3}$
$B^- \to D^*(2007)^0\ D^*(2010)^-\ K^0$	(7.8 ± 2.6)	$\times 10^{-3}$
$B^- \to D^0\ \bar{D}^0\ K^-$	(1.37 ± 0.32)	$\times 10^{-3}$
$B^- \to D^0\ \bar{D}^*(2007)^0\ K^-$	(4.7 ± 1.0)	$\times 10^{-3}$
$B^- \to D^*(2007)^0\ \bar{D}^*(2007)^0\ K^-$	(5.3 ± 1.6)	$\times 10^{-3}$
$B^- \to D^-\ D^*(2007)^+\ K^-$	(1.5 ± 0.4)	$\times 10^{-3}$
$B^- \to D^0\ \omega\ \pi^-$	(4.1 ± 0.9)	$\times 10^{-3}$
$B^- \to D^*(2007)^0\ \omega\ \pi^-$	(4.5 ± 1.2)	$\times 10^{-3}$
$B^- \to D^0\ \pi^+\ \pi^-\ \pi^-$	(1.1 ± 0.4)	$\times 10^{-2}$
$B^- \to D^*(2007)^0\ \pi^+\ \pi^-\ \pi^-\ \pi^0$	(1.8 ± 0.4)	$\times 10^{-2}$
$B^- \to D^{*0}\ 3\pi^-\ 2\pi^+$	(5.7 ± 1.2)	$\times 10^{-3}$
$B^- \to D^*(2010)^+\ \pi^-\ \pi^-\ \pi^0$	(1.5 ± 0.7)	$\times 10^{-2}$
$B^- \to D^*(2010)^+\ \pi^+\ \pi^-\ \pi^-\ \pi^-$	(2.6 ± 0.4)	$\times 10^{-3}$

B.3 D^0 Modes

Channel	Γ_i/Γ	
$D^0 \to K^-\pi^+$	(3.80 ± 0.07)	$\times 10^{-2}$
$D^0 \to K^-\pi^+\pi^0$	(14.1 ± 0.5)	$\times 10^{-2}$
$D^0 \to K^-\pi^+\pi^+\pi^-$	(7.72 ± 0.28)	$\times 10^{-2}$
$D^0 \to K^+K^-$	(3.84 ± 0.10)	$\times 10^{-3}$
$D^0 \to K^0_S K^-\pi^+$	(3.4 ± 0.5)	$\times 10^{-3}$
$D^0 \to K^0_S\pi^0$	(1.14 ± 0.12)	$\times 10^{-2}$
$D^0 \to K^0_S\pi^+\pi^-$	(2.90 ± 0.19)	$\times 10^{-2}$
$D^0 \to K^0_S\pi^+\pi^-\pi^0$	(5.3 ± 0.6)	$\times 10^{-2}$
$D^0 \to K^0_S K^+K^-$	(4.58 ± 0.34)	$\times 10^{-3}$
$D^0 \to K^-\pi^+\pi^+\pi^-\pi^0$	(4.1 ± 0.4)	$\times 10^{-2}$

B.4 D^+ Modes

Channel	Γ_i/Γ	
$D^+ \to K^-\pi^+\pi^+$	(9.51 ± 0.34)	$\times 10^{-2}$
$D^+ \to K^-\pi^+\pi^+\pi^0$	(5.5 ± 2.7)	$\times 10^{-2}$
$D^+ \to K^0_S K^+$	(2.96 ± 0.19)	$\times 10^{-3}$
$D^+ \to K^+K^-\pi^+$	(1.00 ± 0.04)	$\times 10^{-2}$
$D^+ \to K^0_S\pi^+$	(1.47 ± 0.06)	$\times 10^{-2}$
$D^+ \to K^0_S\pi^+\pi^0$	(7.05 ± 0.5)	$\times 10^{-2}$

Channel	Γ_i/Γ	
$D^+ \to K^0_S\pi^+\pi^+\pi^-$	(3.11 ± 0.21)	$\times 10^{-2}$
$D^+ \to K^0_S K^0_S K^+$	(4.7 ± 2.1)	$\times 10^{-3}$
$D^+ \to K^- 3\pi^+\pi^-$	(5.8 ± 0.6)	$\times 10^{-3}$

B.5 $D^*(2007)^0$ Modes

Channel	Γ_i/Γ	
$D^*(2007)^0 \to D^0\pi^0$	(61.9 ± 2.9)	$\times 10^{-2}$
$D^*(2007)^0 \to D^0\gamma$	(38.1 ± 2.9)	$\times 10^{-2}$

B.6 $D^*(2010)^+$ Modes

Channel	Γ_i/Γ	
$D^*(2010)^+ \to D^0\pi^+$	(67.7 ± 0.5)	$\times 10^{-2}$
$D^*(2010)^+ \to D^+\pi^0$	(30.7 ± 0.5)	$\times 10^{-2}$

B.7 D_S^+ Modes

Channel	Γ_i/Γ	
$D_S^+ \to K^+\bar{K}^0$	(4.4 ± 0.9)	$\times 10^{-2}$
$D_S^+ \to \phi\pi^+$	(4.4 ± 0.6)	$\times 10^{-2}$
$D_S^+ \to K^0_S K^*(892)^+$	(5.3 ± 1.3)	$\times 10^{-2}$
$D_S^+ \to \phi\pi^+\pi^0$	(11 ± 5)	$\times 10^{-2}$
$D_S^+ \to K^*(892)^+ K^*(892)^0$	(7.0 ± 2.7)	$\times 10^{-2}$

B.8 D_S^{*+} Modes

Channel	Γ_i/Γ	
$D_S^{*+} \to D_S^+\gamma$	(94.2 ± 0.7)	$\times 10^{-2}$

Bibliography

[BL98] Prof. Dr. Volker Blobel and Prof. Dr. Erich Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. B.G. Teubner, Stuttgart, DESY and Universität Hamburg, 1^{st} edition, 1998.

[Col07] The Belle Collaboration. Software Festa '07. `http://belle.kek.jp/~harat/software2007/Festa-Program.html`, 2007. [Online; accessed 23-September-2009].

[ea95] A. Abashian et al. The belle detector. Technical report, The Belle Collaboration, 1995.

[eaPDGon] C. Amsler et al. (Particle Data Group). Review of particle physics. `http://pdg.lbl.gov`, 2008 and 2009 partial update for the 2010 edition.

[Fei04] Michael Feindt. A neural bayesian estimator for conditional probability densities, 2004.

[Gro01] Belle SVD Group. Belle SVD 2 upgrade technical design report. Technical report, The Belle Collaboration, 2001.

[Hin97] Frank Hinterberger. *Physik der Teilchenbeschleuniger*. Springer-Verlag, Universität Bonn, 1^{st} edition, 1997.

[Per87] Donald H. Perkins. *Introduction to High Energy Physics*. Addison Wesley, University of Oxford, 3^{rd} edition, 1987.

[Rep95] KEK Report. KEK B B-factory design report. `http://www-acc.kek.jp/kekb/publication/KEKB_design_report/KEKBDesignReport.html`, 1995. [Online; accessed 23-September-2009].

[Rep09] KEK Report. KEK B machine parameters. `http://www-acc.kek.jp/kekb/Commissioning/Machineparam2009Jun17.pdf`, 2009. [Online; accessed 23-September-2009].

[T. 09] The CDF Collaboration: T. Aaltonen. Evidence for a narrow near-threshold structure in the $J/\psi\phi$ mass spectrum in B^+ to $J/\psi\phi K^+$ decays. *Physical Review Letters*, 102:242002, 2009.

Danksagung

Für die Aufnahme im Institut für experimentelle Kernphysik, die hervorragende Betreuung und die interessanten Themen, die ich bearbeiten durfte, bedanke ich mich bei Herrn Prof. Dr. Michael Feindt.

Für die Übernahme des Korreferats dieser Arbeit danke ich Herrn Prof. Dr. Günter Quast.

Für die hervorragende Betreuung und das Korrekturlesen dieser Arbeit bedanke ich mich bei Herrn Dr. Thomas Kuhr und Herrn Dr. Michal Kreps.

Für zusätzliches Korrekturlesen bedanke ich mich bei Sabina Geßler, Dr. Andreas Geßler und Sebastian Neubauer.

Für eine entspannte Arbeitsatmosphäre, interessante Diskussionen und stete Hilfsbereitschaft danke ich allen Mitgliedern des Instituts für experimentelle Kernphysik.

Für stete Unterstützung vor und während des Studiums danke ich meinen Eltern Anita und Herbert Zander.