Optimierung der Signalselektion von exklusiv rekonstruierten B^0– und B_s–Mesonzerfällen bei CDF II

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

genehmigte

DISSERTATION

von

Dipl.–Phys. Christian Dörr
aus Gießen

Tag der mündlichen Prüfung: 23.6.2006
Referent: Prof. Dr. Michael Feindt, Institut für Experimentelle Kernphysik
Korreferent: Prof. Dr. Günter Quast, Institut für Experimentelle Kernphysik

Deutsche Zusammenfassung

Einführung

Das Ziel der experimentellen Teilchenphysik ist die Erforschung der fundamentalen Wechselwirkungen zwischen Elementarteilchen. Nach gegenwärtigem Stand der Forschung ist die Welt aus je drei Generationen von Quarks und Leptonen aufgebaut. Die Wechselwirkung der Elementarteilchen wird beschrieben durch die Theorie der elektroschwachen und starken Wechselwirkung im Rahmen des Standardmodells der Teilchenphysik [1, 2, 3, 4]. Die Überprüfung der Vorhersagen des Standardmodells ist eine wichtige Aufgabe der experimentellen Teilchenphysik. In den letzten Jahrzehnten wurden zu diesem Zweck viele Experimente durchgeführt, die die Vorhersagen des Standardmodells bestätigen konnten. All diesen Experimenten ist gemein, daß durch die Kollision von Elektronen oder Protonen bei sehr hohen Energien Teilchen mit einer großen Masse, wie Eichbosonen oder schwere Quarks, erzeugt und untersucht werden. Zur Rekonstruktion dieser Reaktionen dienen Teilchendetektoren, die zum Nachweis der aus den Kollisionen resultierenden Leptonen, Baryonen (gebundene Zustände aus drei Quarks) und Mesonen (gebundene Zustände aus einem Quark und einem Antiquark) verwendet werden. Die Untersuchung dieser Teilchen erlaubt Schlußfolgerungen über die Eigenschaften der der Reaktion zugrundeliegenden Wechselwirkung.
Im Standardmodell der Teilchenphysik unterscheiden sich Eigenzustände der schwachen Wechselwirkung und Masseneigenzustände der Quarks. Dies hat zur Folge, daß Oszillationen zwischen Teilchen– und Antiteilchenzustand von neutralen Mesonen möglich sind. Die Frequenz der Oszillationen hängt vom Massenunterschied Δm zwischen schwerem und leichtem Masseneigenzustand ab. Die 3×3 Transformationsmatrix, die Masseneigenzustände der Quarks in Wechselwirkungseigenzustände rotiert, heißt Cabibbo–Kobayashi–Maskawa– (oder kurz CKM–) Matrix [5, 6].
Solche Meson–Antimesonoszillationen konnten erstmals im Jahr 1956 im $K^0 - \bar{K}^0$ System experimentell nachgewiesen werden [7]. Weitere neutrale Mesonen, bei denen die Möglichkeit zu Teilchen–Antiteilchenoszillationen besteht, sind D^0–, B^0– und B_s–Mesonen. Im Falle des D^0–Mesons erwartet man sehr langsame Oszillationen, so daß diese aufgrund der kurzen Lebensdauer dieses Teilchens schwer nachweisbar sind. Erste Hinweise auf B^0– und B_s–Mesonoszillationen konnten erstmals 1987 am UA1 [8] und ARGUS [9] Experiment beobachtet werden. Eine *zeitaufgelöste* Messung von B–Mesonoszillationen dient der Bestimmung der Oszillationsfrequenz Δm. Die Oszillationsfrequenz Δm_d des B^0–Mesons konnte in den letzten Jahren mit großer Genauigkeit bestimmt werden, der aktuelle Weltmittelwert beträgt $\Delta m_d = 0.507 \pm 0.004\ \mathrm{ps}^{-1}$ [10]. Im Falle des B_s–Mesons erwartet man eine sehr viel höhere und dadurch experimentell schwieriger auflösbare Oszillationsfrequenz Δm_s. Bis vor kurzem war nur eine aus Resultaten der Experimente am LEP e^+e^-–Ringbeschleuniger am CERN in Genf, des SLD Experiments am SLAC e^+e^-–Linearbeschleuniger in Stanford (USA) und des

CDF Experiments am $p\bar{p}$–Ringbeschleuniger Tevatron in Chicago (USA) abgeleitete Untergrenze [1] von $\Delta m_s > 14.4$ ps^{-1} [10] bekannt. Eine erste zeitaufgelöste Messung von B_s–Oszillationen konnte vor kurzem am CDF II Experiment realisiert werden und resultiert in $\Delta m_s = 17.33^{+0.42}_{-0.21} \pm 0.07$ ps^{-1} [11]. Mit Hilfe der Messungen von Δm_d und Δm_s kann erstmals das Verhältnis der CKM–Matrixelemente $\left|V_{ts}/V_{td}\right| = 0.2087^{+0.0047}_{-0.0076}$ mit guter Genauigkeit bestimmt werden. Die Signifikanz des bei CDF II beobachteten Δm_s–Signals beträgt 3.5 σ und die Wahrscheinlichkeit, ein solches Signal durch eine Fluktuation des Untergrunds zu erhalten, beträgt 0.5 %.
Im Rahmen dieser Arbeit wird ein Algorithmus basierend auf der Verwendung neuronaler Netze entwickelt, mit dem Zerfälle von B–Mesonen in exklusive Endzustände[2] effektiv von Untergrundereignissen separiert werden können. Diese Methode wird verwendet, um durch eine effektivere Selektion von Signalereignissen S bei gleichzeitig stärkerer Unterdrückung von Untergrundereignissen B die Signifikanz $\mathcal{S}$ des B_s–Oszillationssignals, gegeben durch [12]

$$\mathcal{S} = \frac{S}{\sqrt{S+B}} \sqrt{\frac{\epsilon D^2}{2}}\, e^{-\sigma_t^2 \Delta m_s^2/2} \tag{1}$$

zu erhöhen. Der Faktor ϵD^2 ist ein Maß für die Leistungsfähigkeit von Taggingalgorithmen, die bestimmen, ob ein rekonstruiertes B–Meson als Meson oder Antimeson erzeugt wurde. Der Faktor $e^{-\sigma_t^2 \Delta m_s^2/2}$ gibt an, wie sehr das Oszillationssignal durch die experimentelle Auflösung σ_t des rekonstruierten Zerfallsvertexes verschmiert wird. Ziel dieser Arbeit ist die Verbesserung des Faktors $\frac{S}{\sqrt{S+B}}$ unter Verwendung neuronaler Netze. Im Rahmen einer aktualisierten Messung von Δm_s am CDF II Experiment wird die daraus resultierende Verbesserung die Signifikanz $\mathcal{S}$ des Δm_s–Signals erhöhen, damit die Wahrscheinlichkeit einer Untergrundfluktuation verringern und so zur endgültigen Etablierung des B_s–Oszillationssignals beitragen.
Über die Selektion von Ereignissen hinaus wird im Rahmen dieser Arbeit eine Methode zur Bestimmung der B–Mesonlebensdauer und der Oszillationsfrequenz Δm für Datensätze entwickelt, deren Lebensdauerverteilungen durch Vorselektionen und die experimentelle Auflösung verformt sind. Diese Methode kann einerseits als unabhängiger Test des Ergebnisses für Δm_s in [11] dienen, andererseits aber auch zur Kalibration von Taggingalgorithmen verwendet werden, die im Rahmen der Messung der CP–Asymmetrie im B_s–System Anwendung finden. Auch die effektive Selektion von exklusiven B–Mesonzerfällen mit Hilfe neuronaler Netze ist nicht auf eine Anwendung im Rahmen einer B_s–Oszillationsanalyse beschränkt, sondern kann in allen Arten von Analysen, für die solche Endzustände relevant sind, Verwendung finden.

[1] In dieser Arbeit wird die Massendifferenz Δm in Einheiten einer Oszillationsfrequenz angegeben, dies impliziert $\hbar = c = 1$.

[2] Bei exklusiven Endzuständen handelt es sich um eindeutig bestimmte Zerfallskanäle, bei denen alle in der Zerfallskette vorkommenden Teilchen vollständig rekonstruiert werden.

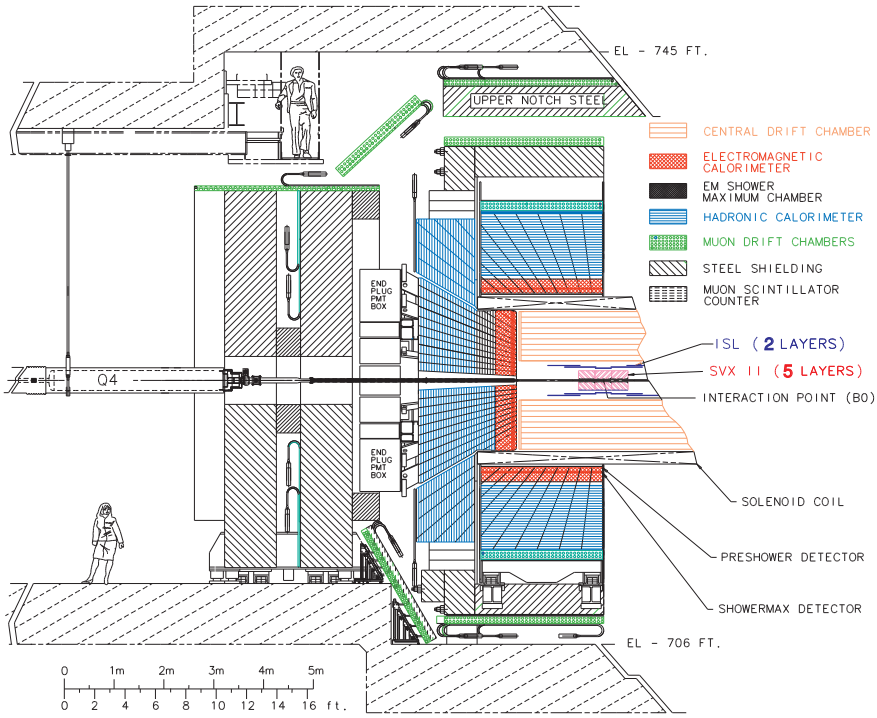


Abbildung 1: Der CDF II Detektor.

Experimenteller Aufbau

Die in dieser Arbeit verwendeten Daten wurden mit Hilfe des CDF II Detektors am Fermi National Accelerator Laboratory in Batavia bei Chicago (USA) gesammelt. Der CDF II Detektor [3] wird von einer internationalen Kollaboration betrieben, an der etwa 500 Physiker und ca. 60 Universitäten und Forschungsinstitute weltweit beteiligt sind. Der CDF II Detektor zeichnet Ereignisse aus Proton–Antiprotonkollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 1.96$ TeV auf und befindet sich an einem der beiden Wechselwirkungspunkte des Tevatrons, eines symmetrischen Kreisbeschleunigers mit einem Radius von 1 km.
Der CDF II Detektor ist ein zylindersymmetrischer Mehrzweckdetektor, der für eine Vielzahl von physikalischen Fragestellungen Verwendung findet. Abbildung 1 zeigt schematisch den Aufbau des Detektors. Die innersten Komponenten des CDF II Detektors sind der Siliziumvertexdetektor und die ihn umgebende Spurkammer COT. Beide Komponenten bilden ein System, das den Nachweis von Spuren geladener Teilchen

[3] CDF steht für “Collider Detector experiment at Fermilab”.

ermöglicht. Außerhalb der Spurkammer erzeugt ein supraleitender Solenoidmagnet ein homogenes Magnetfeld von 1.4 Tesla, das die Spuren der geladenen Teilchen krümmt und so die Messung des Transversalimpulses ermöglicht.
Der Siliziumvertexdetektor besteht aus den Subkomponenten Layer00, SVXII und ISL. Layer00 bildet die innerste Lage von Siliziumsensoren im Abstand $r = 1.6\,cm$ vom Wechselwirkungspunkt. SVXII, das Kernstück des Detektors, besteht aus sechs doppelseitigen Lagen von Siliziumsensoren, die Messungen in allen drei Raumkoordinaten erlauben. Layer00 und SVXII ermöglichen die Rekonstruktion von Zerfallsvertizes kurzlebiger Teilchen mit hoher Präzision. Als Bindeglied zwischen SVXII und der Spurkammer fungiert der aus zwei weiteren Lagen von Siliziumsensoren bestehende ISL Subdetektor. Die Spurkammer selbst besteht aus 96 Lagen von Meßdrähten, die die Messung von Transversalimpulsen von geladenen Teilchen mit hoher Genauigkeit ermöglichen. Zwischen Spurkammer und Solenoidmagnet befindet sich der zur Teilchenidentifikation verwendete Flugzeitdetektor. Außerhalb des Magnetfeldes befinden sich elektromagnetische und hadronische Kalorimeter, die der Energiemessung von Teilchen, bzw. Jets bei hohen Energien dienen. Die Kalorimeter werden von einem System bestehend aus Szintillatoren und Spurkammern zum Nachweis von Myonen umgeben.
Aufgrund der hohen Kollisions– und Wechselwirkungsrate bei CDF II ist es weder möglich noch wünschenswert, alle resultierenden Ereignisse auf Speichermedien aufzuzeichnen. Nur wenige dieser Ereignisse sind von physikalischem Interesse, daher findet bei CDF II ein aufwendiges, mehrstufiges Filtersystem (*Trigger*) zur Vorselektion von potentiell interessanten Ereignissen Anwendung. Der für die vorliegende Arbeit verwendete Datensatz basiert vollständig auf Ereignissen, die mit Hilfe des Zweispurtriggers vorselektiert wurden. Der dem Zweispurtrigger zugrundeliegende Gedanke ist, daß Spuren, die aus dem Zerfall eines Teilchen mit einer vergleichsweise langen Lebensdauer resultieren, wie z. B. eines B–Mesons, durch eine Verschiebung ihres Stoßparameters in Bezug auf den Wechselwirkungspunkt identifiziert werden können. Abbildung 2 veranschaulicht den Zusammenhang zwischen dem Stoßparameter der Spuren und der Lebensdauer des Teilchens. Der Zweispurtrigger selektiert Ereignisse, die mindestens zwei vom Wechselwirkungspunkt verschobene Spuren mit einem Transversalimpuls von $p_t > 2$ GeV[4] beinhalten.

[4]In dieser Arbeit werden Energie, Impuls und invariante Masse in der Einheit GeV angegeben. Ein GeV entspricht $1.602\,176\,462(63) \cdot 10^{-10}$ Joule, die Umrechnung von Impulsen (Massen) in die Einheiten $\frac{kg\,m}{s}$ (kg) erfolgt mit der Division durch c (c^2), wobei $c = 299\,792\,458\,\frac{m}{s}$.

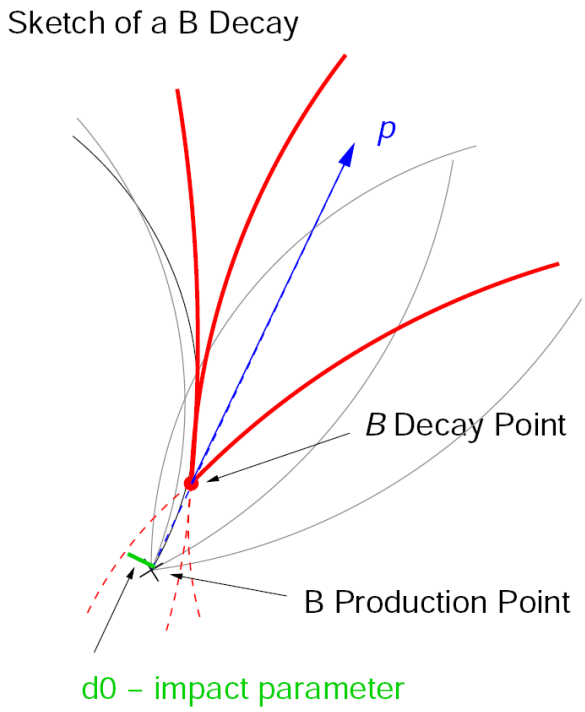


Abbildung 2: Veranschaulichung der Funktionsweise des Zweispurtriggers. Die Lebensdauer der B–Mesonen führt zu einer Verschiebung von Zerfallsort (*decay point*) und Produktionsort (*production point*), Spuren aus diesem Zerfall zeichnen sich durch einen vom Produktionsort verschobenen Stoßparameter d0 aus.

Rekonstruktion von B–Mesonen

Die in dieser Arbeit verwendeten Daten wurden am CDF II Experiment im Zeitraum von Februar 2002 bis Dezember 2005 mit Hilfe des Zweispurtriggers gesammelt und entsprechen einer integrierten Luminosität von 765 pb^{-1}.
Trotz der Anreicherung des Datensatzes mit langlebigen Teilchen durch den Zweispurtrigger enthält nur ein winziger Bruchteil der Ereignisse tatsächlich den Zerfall eines B–Mesons. Statt zweier aus dem Zerfall eines langlebigen Teilchens stammender Spuren erfüllt oft lediglich eine zufällige Kombination von zwei aus Untergrundereignissen resultierenden Spuren das Triggerkriterium und wird selektiert. Es bedarf eines ausgereiften und effektiven Rekonstruktionsalgorithmus, der es ermöglicht, aus der Vielzahl von Ereignissen diejenigen zu identifizieren, die den Zerfall eines B–Mesons in einen exklusiven Endzustand enthalten.

Die Grundlage zur Rekonstruktion exklusiver Endzustände bildet die präzise Messung von Trajektorien geladener Teilchen durch Spurrekonstruktionsalgorithmen. Da bei der Rekonstruktion von B–Mesonen sowohl eine sehr genaue Bestimmung des Zerfallsvertexes als auch eine präzise Bestimmung des transversalen Impulses von Interesse sind, werden Spuren verwendet, die sowohl im Siliziumvertexdetektor als auch in der Spurkammer nachgewiesen wurden und so eine genaue Bestimmung aller Spurparameter erlauben. Der für diese Arbeit verwendete Spurrekonstruktionsalgorithmus basiert auf dem Kalmanfitter [13] und kombiniert die Information beider Detektoren durch Extrapolation von Spuren der Spurkammer in den Siliziumvertexdetektor. Die Rekonstruktion des Zerfalls eines B–Mesons in einen exklusiven Endzustand erfolgt durch die sukzessive Rekonstruktion der Zerfallsvertizes aller in der Zerfallskette auftretenden Zwischenzustände. So beginnt z.B. die Rekonstruktion des Zerfalls $B^0 \to D\pi, D \to K\pi\pi$ mit der Rekonstruktion des Zerfalls $D \to K\pi\pi$, indem alle Kombinationen von Spuren durch einen Vertexfit auf die Hypothese überprüft werden, ob die drei Spuren ihren gemeinsamen Ursprung an einem Punkt haben und ob die invariante Masse der Spurkombination der Ruhemasse des D–Mesons entspricht. Als nächster Schritt werden alle D–Mesonkandidaten mit den verbliebenen Spuren kombiniert und auf die Hypothese überprüft, ob D–Meson und vierte Spur mit dem Zerfall eines B^0–Mesons kompatibel sind. Ereignisse, die nicht mit dieser Hypothese vereinbar sind, können als Untergrundereignisse klassifiziert und verworfen werden.
Der auf diese Weise vorselektierte und mit B–Mesonen angereicherte Datensatz dient als Grundlage für den im Rahmen dieser Arbeit entwickelten Algorithmus zur optimierten Selektion von B–Mesonzerfällen mit Hilfe neuronaler Netze.

Signaloptimierung mit Neuronalen Netzen

Die Signifikanz $\mathcal{S}$, mit der die B_s–Oszillationsfrequenz Δm_s gemessen werden kann, wird von verschiedenen experimentellen Faktoren limitiert (siehe Gleichung 1). Einer dieser Faktoren ist die Signalsignifikanz $\frac{S}{\sqrt{S+B}}$, deren Optimierung der Fokus der vorliegenden Arbeit gilt. Sie wird bestimmt durch die Anzahl von Signalereignissen S, die bei einem Niveau von B Untergrundereignissen beobachtet werden. Das bisher bei CDF II verwendete Verfahren zur Optimierung der Signalsignifikanz $\frac{S}{\sqrt{S+B}}$ basiert auf einer sequentiellen Schnittoptimierung und nutzt Informationen zur Separation von Signal– und Untergrundereignissen in nicht optimaler Weise. Die Methode verwendet zur Separation nur eine begrenzte Zahl von Kenngrößen (Variablen) und vernachlässigt insbesondere deren Korrelationen vollständig. Der im Rahmen dieser Arbeit unter Verwendung neuronaler Netzwerke entwickelte Optimierungsalgorithmus ist dagegen in der Lage, alle zur Charakterisierung von B–Mesonzerfällen zur Verfügung stehenden

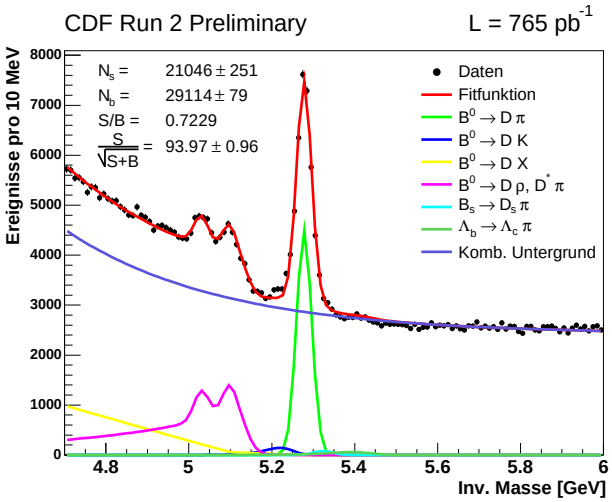


Abbildung 3: Zusammensetzung des Datensatzes, der für das Netzwerktraining im Kanal $B^0 \rightarrow D\pi, D \rightarrow K\pi\pi$ verwendet wird. Muster für Untergrundereignisse für das Netzwerktraining werden aus dem Massenbereich von 5.4 bis 5.6 GeV des oberen Seitenbands entnommen.

Variablen zu verwenden und diese unter Berücksichtigung aller Korrelationen optimal zu kombinieren. Zur Optimierung der Kennzahl $\frac{S}{\sqrt{S+B}}$ wird ein neuronales Netzwerk trainiert, das B–Mesonsignalereignisse von Untergrundereignissen zu trennen lernt. Das Ziel ist, für jedes Ereignis im zur Verfügung stehenden Datensatz die Wahrscheinlichkeit, daß es sich um den Zerfall eines B–Meson handelt, zu berechnen und den Schnitt auf diese Wahrscheinlichkeit, der $\frac{S}{\sqrt{S+B}}$ maximiert, zu ermitteln. Es handelt sich dabei um ein binäres Klassifizierungsproblem, wofür man Muster für Signal– und Untergrundereignisse benötigt, um das Netzwerk zu trainieren. Als Muster für Signalereignisse werden mit Hilfe einer PYTHIA–Simulation [14] erzeugte B–Mesonzerfälle verwendet. Muster für Untergrundereignisse können nicht mit Hilfe einer Simulation erzeugt werden, da die komplexe Zusammensetzung des an einem Hadronenbeschleuniger beobachteten Untergrunds bisher nicht vollständig durch Modelle beschrieben werden kann. Ein alternativer Ansatz besteht in der Verwendung von Untergrundereignissen aus realen Daten, wie im folgenden dargelegt wird. Abbildung 3 zeigt die Verteilung der invarianten B–Mesonmasse im Endzustand $B^0 \rightarrow D\pi, D \rightarrow K\pi\pi$. Im Bereich von 5.2 bis 5.35 GeV ist das Resonanzsignal des Zerfalls sichtbar. Im Bereich niedrigerer invarianter Masse sind Strukturen sichtbar, die unvollständig rekonstruierten B^0–Zerfällen zugeordnet werden können. Im Bereich über 5.4 GeV wird das Massenspektrum vollständig von kombinatorischem Untergrund, resultierend aus zufälligen Kombinationen von Spuren, dominiert. Ziel der Signaloptimierung ist, den Beitrag die-

ses kombinatorischen Untergrunds in der Signalregion weitgehend zu eliminieren und dabei weiterhin möglichst effektiv Signalereignisse zu selektieren. Um dies zu erreichen, wird ein neuronales Netzwerk zur Separation von B–Mesonsignalereignissen von kombinatorischem Untergrund trainiert, wobei Ereignisse aus realen Daten im Massenbereich von 5.4 bis 5.6 GeV als Muster für kombinatorische Untergrundereignisse dienen. Basierend auf dem Training werden alle Ereignisse des Datensatzes klassifiziert und die Signalsignifikanz $\frac{S}{\sqrt{S+B}}$ durch Selektion des optimalen Schnittes auf die Ausgabe des Netzwerks maximiert.
Diese Strategie wird im Rahmen dieser Arbeit am Beispiel des Zerfalls $B^0 \to D\pi, D \to K\pi\pi$ entwickelt und mögliche systematische Effekte werden eingehend untersucht. Im Vergleich zur bisher bei CDF II verwendeten sequentiellen Schnittoptimierung kann die Signalsignifikanz $\frac{S}{\sqrt{S+B}}$ durch die Verwendung eines neuronalen Netzwerks in diesem Zerfallskanal um 11 % verbessert werden. Wollte man den gleichen Wert mit einer schnittbasierten Selektion erreichen, benötigte man dazu 24 % mehr an Daten [5]. Abbildung 4 vergleicht die schnittbasierte mit der aus der Verwendung eines neuronalen Netzwerks resultierenden Selektion. Beide Selektionen wurden auf maximale Signalsignifikanz $\frac{S}{\sqrt{S+B}}$ hin optimiert. Man kann der Abbildung entnehmen, daß das Netzwerk auf dem gleichen Datensatz bei einem ähnlichen Untergrundniveau sehr viel effektiver Signalereignisse selektiert.
Die Optimierungsstrategie kann direkt auf weitere exklusive B–Mesonzerfälle angewandt werden. Um optimale Ergebnisse zu erzielen, muß jedoch jeder einzelne Endzustand auf besondere Charakteristika, wie Winkelkorrelationen zwischen den Tochterteilchen oder das Vorhandensein von Subresonanzen, untersucht werden, um dem Netzwerk alle relevanten Informationen zur Separation von Signal und Untergrund zur Verfügung zu stellen. Als weiteres Beispiel mit direkter Relevanz für eine Messung von Δm_s wird die Selektionsmethode auf den Zerfallskanal $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ angewandt. Im Vergleich zur mit Schnitten erzielten Signalsignifikanz kann unter Verwendung eines neuronalen Netzwerks der maximale Wert für $\frac{S}{\sqrt{S+B}}$ um 16 % verbessert werden. Wollte man den gleichen Wert mit einer schnittbasierten Selektion erreichen, benötigte man dazu 35 % mehr an Daten. Die verbesserte Signalselektion läßt, unter Berücksichtigung der Ortsauflösung σ_t der selektierten Ereignisse (siehe Gleichung 1), eine Erhöhung der Signifikanz der Messung von Δm_s bei CDF II im Zerfallskanal $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ um 14 % erwarten.

[5] In diesen Wert geht die Annahme ein, daß das Verhältnis von Signal– zu Untergrundereignissen S/B der schnittbasierten Selektion konstant bleibt, wenn weitere Daten hinzugefügt werden.

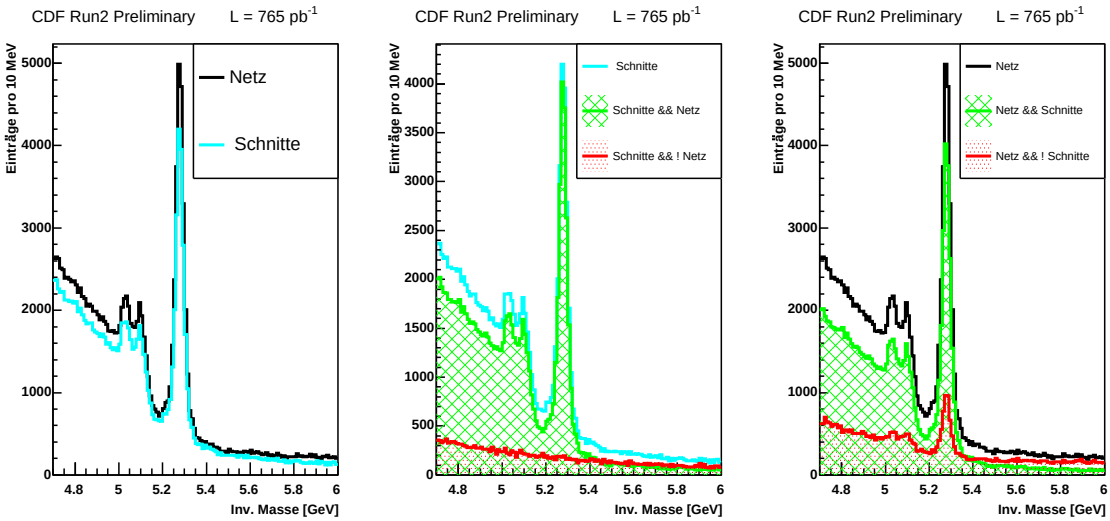


Abbildung 4: Vergleich der durch Schnitte/neuronales Netz selektierten Ereignisse. Die linke Abbildung zeigt die Verteilung der invarianten Massen von durch Schnitte (hellblau, "Schnitte") bzw. neuronales Netz (schwarz, "Netz") selektierten B–Mesonkandidaten. Die mittlere und rechte Abbildung verdeutlichen die Gemeinsamkeiten (grüne Verteilung) und Unterschiede (rote Verteilung) beider Selektionen. Insbesondere kann man der rechten Abbildung entnehmen, daß das neuronale Netz zusätzliche Signalereignisse selektiert, die durch Anwendung der Schnittselektion verworfen worden wären.

Methode zur Bestimmung von Δm

Im Rahmen dieser Arbeit wird eine Maximum–Likelihood–Methode zur direkten Bestimmung der B–Mesonlebensdauer τ und der Oszillationsfrequenz Δm implementiert um zu demonstrieren, daß sich die Ereignisselektion mit Hilfe eines neuronalen Netzes in einer Messung von Δm verwenden läßt. Die Methode basiert auf der Beschreibung der Verteilungen der invarianten Massen m und Lebensdauern t von B–Mesonsignal– (S) und Untergrundereignissen (B). Für eine Bestimmung der Oszillationsfrequenz muß darüber hinaus die Information, ob ein Meson als (Anti–)Teilchen erzeugt wurde und zum Zeitpunkt t als (Anti–)Teilchen zerfallen ist, vorliegen. Diese Information wird durch Taggingalgorithmen ermittelt und ist in der Variablen ξ enthalten. Die Wahrscheinlichkeitsdichte $P(m, t, \xi, \sigma_t)$, die diese Informationen für Signal– und Un-

tergrundereignisse berücksichtigt, hat die Form

$$P(m,t,\xi,\sigma_t) = (1-f_B)\cdot P_S(m)\cdot P_S(t,\xi,\sigma_t) \;+\; f_B\cdot P_B(m)\cdot P_B(t)$$

$P_{S,B}(m)$ sind dabei Parametrisierungen der Verteilungen der invarianten Massen von Signal– und Untergrundereignissen. Der Parameter f_B gewichtet die relativen Anteile von Signal– und Untergrundereignissen. $P_B(t)$ parametrisiert die Lebensdauerverteilung von Untergrundereignissen und $P_S(t,\xi,\sigma_t)$ enthält die Lebensdauerverteilung von B–Mesonsignalereignissen unter Berücksichtigung der Teilchen–Antiteilchenoszillationen. $P_S(t,\xi,\sigma_t)$ hat die Form

$$P_S(t,\xi,\sigma_t) = \frac{1}{N_S(t,\xi,\sigma_t)}\left(\frac{1+\xi D\cos(\Delta m\, t)}{1+|\xi|}\,\frac{1}{\tau}e^{-\frac{t}{\tau}}\right)\otimes\mathcal{G}(t-t',\sigma_t)\cdot\epsilon(t)$$

und berücksichtigt die Verschmierung des Oszillationssignals durch die Faltung mit einer Auflösungsfunktion $\mathcal{G}(t-t',\sigma_t)$ und die limitierte Akzeptanz in Form der Effizienzfunktion $\epsilon(t)$. Der Faktor $N_S(t,\xi,\sigma_t)$ dient der Normierung der Wahrscheinlichkeitsdichte und D ist ein Maß für die Reinheit (*Dilution*) der Entscheidung des Taggingalgorithmus.
Durch die Minimierung der negativen Log Likelihood Funktion

$$\mathcal{L} = -2\sum_i \ln\Big[\big(1-f_B\big)\cdot P_S(t_i,\sigma_{t,i})\cdot P_S(m_i) + f_B\cdot P_B(t_i)\cdot P_B(m_i)\Big]$$

für die Bestimmung der Lebensdauer der B–Mesonen bzw.

$$\mathcal{L} = -2\sum_i \ln\Big[\big(1-f_B\big)\cdot P_S(t_i,\xi_i,\sigma_{t,i})\cdot P_S(m_i) + f_B\cdot P_B(t_i)\cdot P_B(m_i)\Big]$$

zur Bestimmung der Oszillationsfrequenz Δm können im Rahmen eines Unbinned Maximum Likelihood Fits unter Verwendung aller selektierter Ereignisse i mit ihrer jeweils unterschiedlichen Ortsauflösung $\sigma_{t,i}$ die Lebensdauer der B–Mesonen, bzw. Δm bestimmt werden.
Die Anwendung der Methode auf einen durch das neuronale Netz selektierten Datensatz demonstriert, daß es möglich ist, die Verschiebung in der Verteilung der B–Mesonlebensdauer durch Trigger– und Netzwerkselektion mit Hilfe einer Simulation zu beschreiben. Die durch ein neuronales Netzwerk im Zerfallskanal $B^0 \to D\pi, D \to K\pi\pi$ selektierten Ereignisse konnten verwendet werden, um die Lebensdauer τ des B^0–Mesons und dessen Oszillationsfrequenz Δm_d

$$c\tau(B^0) = 460.8 \pm 4.5 \text{ (stat.) } \mu m$$
$$\Delta m_d = 0.507 \;\pm\; 0.069 \text{ (stat.) } ps^{-1}$$

in guter Übereinstimmung mit den Weltmittelwerten $c\tau(B^0) = 459.3 \pm 2.0\,\mu m$ [15] und $\Delta m_d = 0.507 \pm 0.004$ ps^{-1} [10] zu bestimmen.
Die Implementierung der Methode zur Bestimmung von Δm kann als unabhängiger Test des bisher von CDF II veröffentlichten Ergebnisses für Δm_s dienen. In die Messung von Δm_s bei CDF II gehen insgesamt fünf exklusive hadronische Endzustände des B_s Mesons ein. Sowohl zur Überprüfung des Ergebnisses als auch für eine Verbesserung durch die Verwendung neuronaler Netze zur Ereignisselektion müssen also mehrere Zerfallskanäle in die Analyse einbezogen werden.
Die Verwendung der durch das neuronale Netzwerk selektierten Ereignisse im Zerfallskanal $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ für einen Amplitudenscan [12] von Δm_s zeigt, daß die auf der Verwendung des neuronalen Netzwerks basierte Selektion eine deutliche Verbesserung der Signifikanz der Messung von Δm_s erwarten läßt. Der Amplitudenscan mißt den Anteil einer bestimmten Oszillationsfrequenz indem die Amplitude A des Oszillationssignals als freier Parameter im Fit für jeweils fest Werte der Oszillationsfrequenz Δm_s angepaßt wird. Idealerweise nimmt die Amplitude für die im Datensatz enthaltene Oszillationsfrequenz den Wert eins an und für alle anderen Frequenzen den Wert null. Abbildung 5 zeigt, daß die Sensitivität für eine Messung von Δm_s in dem Zerfallskanal $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ bei der Verwendung des aus 765 pb^{-1} bestehenden Datensatzes von $\simeq 17$ ps^{-1} auf $\simeq 18.5$ ps^{-1} erhöht wird. Die Sensitivität wird durch einen Amplitudenscan mit zufälliger Taggingentscheidung ermittelt und ist gegeben durch die Frequenz, bei der der Fehler auf die Amplitude $1.65\,\sigma_A$ (95 % Konfidenzintervall) die Linie $A = 1$ schneidet.
Die Verwendung des durch das Netzwerk selektieren Datensatzes in einem Unbinned Maximum Likelihood Fit von Δm_s ergibt für diesen Zerfallskanal das vorläufige Ergebnis

$$\Delta m_s = 17.64 \pm 0.22 \text{ (stat.) } ps^{-1}$$

Für die aus dem Amplitudenscan abgeleitete Sensitivität und für das Ergebnis von Δm_s wurde kein systematischer Fehler ermittelt, nur die statistische Unsicherheit der Amplituden und von Δm_s ist angegeben, so daß es sich nicht um eine Messung der Sensitivität oder Δm_s handelt. Sowohl Amplitudenscan als auch der Fit von Δm_s dienen lediglich dem Vergleich der schnittbasierten und der auf dem neuronalen Netz basierten Selektionen. Für eine Messung von Δm_s unter Verwendung der Selektion durch neuronale Netzwerke müssen weitere exklusive Zerfallskanäle des B_s–Mesons in die Analyse einbezogen werden. Die Steigerung der Sensitivität der Messung durch die Verwendung der auf dem neuronalen Netz basierenden Selektion läßt eine deutlich erhöhte Signifikanz des bei CDF II beobachteten B_s–Oszillationssignals erwarten, sobald alle Zerfallskanäle berücksichtigt werden.

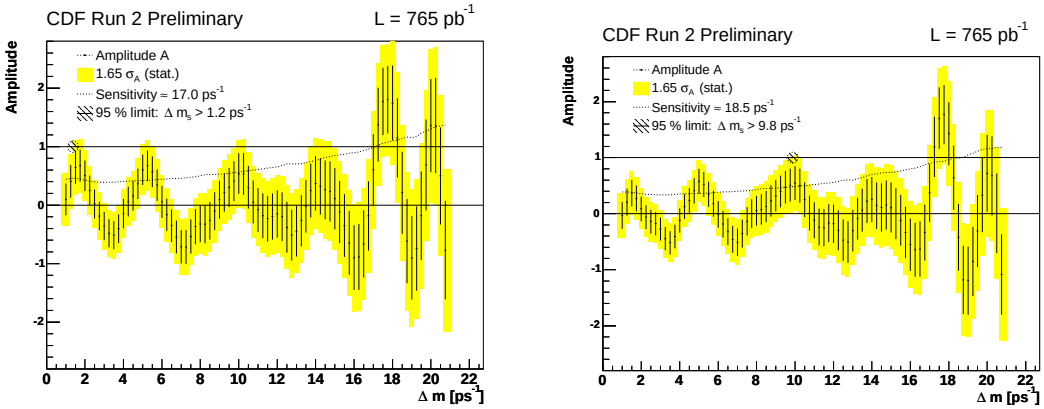


Abbildung 5: Ergebnis des Amplitudenscans für Δm_s im Vergleich für Ereignisse, die mit der herkömmlichen Schnittmethode (links) bzw. durch ein neuronales Netz (rechts) im Zerfallskanal $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ selektiert wurden. Die Sensitivität ("Sensitivity"), gegeben durch den Schnittpunkt der gepunkteten Linie mit der Geraden Amplitude = 1, wird von $\simeq 17$ ps^{-1} auf $\simeq 18.5$ ps^{-1} erhöht.

Zusammenfassung und Ausblick

Im Rahmen der vorliegenden Arbeit wurde eine Methode zur Optimierung der Signalsignifikanz $\frac{S}{\sqrt{S+B}}$ für Zerfälle von B–Mesonen in exklusive Endzustände unter Verwendung neuronaler Netze entwickelt. Im Vergleich zu der bisher bei CDF II angewandten Methode der sequentiellen Schnittoptimierung kombiniert ein neuronales Netz die zur Verfügung stehende Information unter Berücksichtigung von Korrelationen optimal. Die Methode wurde am Beispiel des Zerfallskanals $B^0 \to D\pi, D \to K\pi\pi$ entwickelt und validiert. Im Zerfallskanal $B^0 \to D\pi, D \to K\pi\pi$ verbessert die auf dem neuronalen Netz basierte Ereignisselektion die Signalsignifikanz $\frac{S}{\sqrt{S+B}}$ um 11 %. Um das gleiche Ergebnis mit Hilfe der schnittbasierten Selektion zu erzielen, benötigte man dafür einen um 24 % größeren Datensatz. Die Methode zur Bestimmung der B–Mesonlebensdauer und der Oszillationsfrequenz Δm konnte die durch das neuronale Netz selektierte Ereignisse verwenden, um im Zerfallskanal $B^0 \to D\pi, D \to K\pi\pi$ mit dem Weltmittelwert konsistente Ergebnisse für Lebensdauer und Oszillationsfrequenz zu bestimmen.
Die auf der Verwendung eines neuronalen Netzwerks basierende Selektionsmethode wurde ebenfalls auf den für eine Bestimmung von Δm_s relevanten Zerfallskanal $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ angewandt und konnte die Signalsignifikanz $\frac{S}{\sqrt{S+B}}$ um 16 % erhöhen. Um das gleiche Ergebnis mit Hilfe der schnittbasierten Selektion zu erzielen, benötigte man einen um 35 % größeren Datensatz. Die effekti-

vere Ereignisselektion läßt eine Verbesserung der Signifikanz der Δm_s–Messung bei CDF II um 14 % in diesem Zerfallskanal erwarten. Der Vergleich der im Zerfallskanal $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ durch Schnitte und Netzwerk selektierten Ereignisse in einem Unbinned Maximum Likelihood Fit zur direkten Bestimmung oder in einer Amplitudenscanmethode zur indirekten Bestimmung von Δm_s bestätigt diese Erwartung. Die Methode zur Signaloptimierung unter Verwendung neuronaler Netze kann leicht auf weitere in der Messung von Δm_s verwendete B_s–Zerfallskanäle erweitert werden und läßt, wenn angewandt auf komplexere Zerfälle wie $B_s \to D_s\pi\pi\pi$, eine noch weitaus größere Verbesserung der Signalsignifikanz erwarten. Die verbesserte Signalselektion wird im Rahmen eines aktualisierten Ergebnisses wesentlich zur endgültigen Etablierung des B_s–Oszillationssignals bei CDF II beitragen.
Über eine Verwendung im Rahmen einer Oszillationsanalyse hinaus kann der Algorithmus zur Signalselektion in allen Arten von Analysen verwandt werden, für die Zerfälle von B–Mesonen in exklusive Endzustände relevant sind. Die Methode zur Bestimmung von Δm_s kann in Zukunft zur Kalibration von Taggingalgorithmen zur Verwendung bei einer Messung der CP–Asymmetrie im B_s–System verwendet werden.

Optimization of the signal selection of exclusively reconstructed decays of B^0 and B_s Mesons at CDF-II

Christian Dörr

Ph.D. Thesis
Faculty of Physics
University of Karlsruhe (TH)

Supervisor: Prof. Dr. M. Feindt, Institut für Experimentelle Kernphysik
University of Karlsruhe (TH)

Contents

Introduction **15**

1 Theoretical Overview **19**

1.1 Overview 19

1.2 The CKM Matrix 20

1.3 The Unitarity Triangle 21

1.4 Constraining the Unitarity Triangle 21

1.4.1 $|V_{ub}/V_{cb}|$ 23

1.4.2 $\sin 2\beta$ 23

1.4.3 ϵ_K 23

1.4.4 Δm_d and Δm_s 23

2 Experimental Setup **27**

2.1 The Tevatron 27

2.1.1 The Accelerator Chain 27

2.1.2 Performance 29

2.2 The CDF II Detector 31

2.2.1 Overview 31

2.2.2 The Tracking System 31

2.2.3 Particle Identification 33

2.2.4 Calorimeters 34
2.2.5 Muon Systems 34
2.3 The CDF II Trigger System 35
2.3.1 Overview 35
2.3.2 Level 1 Trigger 35
2.3.3 Level 2 Trigger 35
2.3.4 Level 3 Trigger 36

3 Mathematical Methods 39
3.1 Parameter Estimation 39
3.1.1 The Maximum Likelihood Method 39
3.1.2 The Maximum Likelihood Method for Binned Data 41
3.2 Neural Networks 42
3.2.1 Network Training 44
3.2.2 Preprocessing of Variables 44
3.2.3 Regularization 45
3.2.4 The Bayesian Approach 45

4 Dataset and Reconstruction 47
4.1 Dataset Selection 47
4.1.1 The Two Track Trigger 47
4.1.2 "Good Run" Selection 48
4.2 Event reconstruction 50
4.2.1 Tracking 50
4.2.2 Momentum Scale Calibration and Track Refitting 51
4.2.3 Trigger Confirmation 52
4.2.4 B Meson Reconstruction 53

4.2.5 B Stntuples 55

5 Outline of the Mixing Analysis 59

5.1 Overview 59

5.2 Understanding the Invariant Mass Spectrum 61

5.2.1 The Mass Spectrum for $B^0 \to D\pi, D \to K\pi\pi$ 63

5.2.2 The Mass Spectrum for $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ 68

5.3 Measuring the B Meson Lifetime 69

5.3.1 Introduction 69

5.3.2 Probability Density Function in Lifetime Space 71

5.4 Flavor Tagging 72

5.5 Extracting Δm_d and Δm_s 74

5.5.1 Amplitude Scan 75

5.6 Sensitivity for Δm_s 76

5.6.1 Tagging Performance 77

5.6.2 Proper Decay Time Resolution 77

5.6.3 Signal Significance 78

6 Signal Optimization with Neural Networks 81

6.1 Motivation 81

6.2 Strategy 82

6.3 Obtaining Simulated Events 85

6.4 Selection of Training Variables 87

6.5 Results for $B^0 \to D\pi, D \to K\pi\pi$ 91

6.6 Validation and Checks 97

6.6.1 Selection of the Sideband Region 97

6.6.2 Bias Check 97
6.6.3 Mass Dependence of Variables 102
6.7 Results for $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ 107
6.8 Summary 112

7 Mixing Fit 117
7.1 Overview 117
7.2 Lifetime Fit with a Neural Network Selection 119
7.2.1 Likelihood Function 119
7.2.2 Determining the Efficiency Curve 120
7.2.3 Results of the Lifetime Fit 122
7.3 Fit of Δm_d 124
7.4 Amplitude Scan and Fit of Δm_s 126

8 Conclusion and Outlook 129

A Definition of Variables 131

B Details of the Variable Significance Table 134

C Derivation of Analytic Expressions 136
C.1 Analytical Results of the Normalization Integrals 138
C.2 Numerical Tests of the Analytic Implementation 140

D Comparisons between Data and Simulated Events 144
D.1 Overview 144
D.2 Comparison of Distributions after Cuts 145
D.3 Comparison of Correlations in Data and Simulated Events 155
D.4 Comparison of Data and Simulated Events after a Network Selection . 160

E Compilation of Fit Templates 163

List of Figures

1 Der CDF II Detektor. III

2 Veranschaulichung der Funktionsweise des Zweispurtriggers. V

3 Zusammensetzung des Datensatzes $B^0 \to D\pi, D \to K\pi\pi$ VII

4 Vergleich der durch Schnitte/neuronales Netz selektierten Ereignisse. . IX

5 Amplitudenscan für Δm_s XII

6 Amplitude scans for Δm_s 16

1.1 Three families of fermions 20

1.2 The unitarity triangle 21

1.3 Experimental constraints on the unitarity triangle. 22

1.4 Diagrams giving rise to B mixing 24

2.1 The Fermilab accelerator complex for Run 2. 28

2.2 Integrated Tevatron Run 2 luminosity. 30

2.3 Tevatron Run 2 peak luminosity. 30

2.4 The CDF II detector. 31

2.5 η coverage of the SVX and COT detectors. 32

2.6 Location of the TOF system in the CDF II detector. 34

2.7 Separation power of the particle ID system 34

2.8 Data-flow in the CDF II trigger and data acquisition system. 37

2.9 The first two trigger levels and the involved detector elements. 37
2.10 Definition of the track impact parameter. 37
2.11 SVT impact parameter resolution. 37

3.1 Schematic display of a neural network 43
3.2 Representation of a neural network node. 43

4.1 Effective prescale factors 49
4.2 The removal of the p_t dependence of the J/ψ mass 52
4.3 Topology of the decay $B^0 \to D\pi, D \to K\pi\pi$ 55
4.4 Example of an event in B Stntuple format. 57

5.1 Typical shape of the invariant mass spectrum of a B meson 62
5.2 Invariant mass distribution for reflections, part1 64
5.3 Invariant mass distribution for reflections, part2 66
5.4 Fit of $B^0 \to D\pi, D \to K\pi\pi$ mass spectrum 68
5.5 Fit of $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ mass spectrum 70
5.6 Visualization of different flavor tagging algorithms. 73
5.7 Amplitude scan for Δm_s 76
5.8 Comparison of the time dependent asymmetries 79

6.1 Composition of the training sample 84
6.2 Procedure for the classification of a data sample 84
6.3 Examples of heavy-flavor production diagrams 86
6.4 Correlation matrix of all training variables 88
6.5 Performance of the network when removing variables 90
6.6 Correlation matrix of the training variables 93

6.7 Quality plots from the NeuroBayes® training 93
6.8 Summary of results derived from mass fits 94
6.9 Mass distribution for $B^0 \to D\pi$ after cuts and network 95
6.10 Comparison of events selected by cuts and by the neural network. . . . 96
6.11 Choice of a sideband region 98
6.12 Check of network bias, counted background events 100
6.13 Check of network bias, signal events and B^0 mass 101
6.14 Mass dependence of $\Delta R(D, \pi_B)$ and $L_{xy}/\sigma_{L_{xy}}(B)$ 103
6.15 Mass dependence of $p_t(\pi_B)$ 103
6.16 Network output against invariant mass of the B candidates 104
6.17 Network output for sideband training 105
6.18 Mass distribution after sideband training 106
6.19 Upper sideband for cuts and network selection 108
6.20 Correlation matrix for $B_s \to D_s\pi$ variables 111
6.21 Quality plots for $B_s \to D_s\pi$ training 112
6.22 Summary of results for $B_s \to D_s\pi$ 113
6.23 Mass distribution $B_s \to D_s\pi$ for cuts and network selection 114
6.24 Δm_s significance for different network cuts 115

7.1 Flow chart summarizing the fit procedure 118
7.2 Comparison data/simulation for nn_{out} and ct of the B^0 121
7.3 Fits of the parametrized $\epsilon(t)$ 122
7.4 Projections in lifetime space 123
7.5 Projections in mass space 123
7.6 Likelihood for Δm_d 125

7.7 Amplitude scan for Δm_s 127

7.8 Likelihood for Δm_s 128

A.1 Definition of the lifetime–signed impact parameter 133

C.1 Comparison of analytical/numerical integrals, lifetime term 142

C.2 Comparison of analytical/numerical integrals, mixing term 143

D.1 Comparison of data/simulation of $|d_0(B)|$ and $\chi^2_{r\phi}(B)$ 145

D.2 Comparison of data/simulation of $L_{xy}/\sigma_{L_{xy}}(B)$ and $L_{xy}(B)$ 146

D.3 Comparison of data/simulation of $p_t(B)$ and $p_t(\pi_B)$ 146

D.4 Comparison of data/simulation of $\chi^2_{3D}(B)$ and $\chi^2_{3D}(D)$ 147

D.5 Comparison of data/simulation of $|d_0(D)|$ and $\chi^2_{r\phi}(D)$ 147

D.6 Comparison of data/simulation of $L_{xy}/\sigma_{L_{xy}}(B)$ and $L_{xy}(D)$ 148

D.7 Comparison of data/simulation of the invariant $K-\pi$ masses 148

D.8 Comparison of data/simulation of $L_{xy}(B \to D)$ and $\Delta R(D, \pi_B)$ 149

D.9 Comparison of data/simulation of the lifetime–signed $d_0(\pi_B)$ 149

D.10 Comparison of data/simulation of the lifetime–signed $d_0(K)$ 150

D.11 Comparison of data/simulation of the lifetime–signed $d_0(\pi^1_D)$ 150

D.12 Comparison of data/simulation of the lifetime–signed $d_0(\pi^2_D)$ 151

D.13 Comparison of data/simulation of $\sphericalangle(\vec{p}(B), \vec{p}(\pi_B))_{\mathrm{cms}B}$ and $p_t(K)$ . . . 151

D.14 Comparison of data/simulation of $p_t(\pi^1_D)$ and $p_t(\pi^2_D)$ 152

D.15 Comparison of data/simulation of $|d_0(\pi_B)|$ and $|d_0(K)|$ 152

D.16 Comparison of data/simulation of $|d_0(\pi^1_D)|$ and $|d_0(\pi^2_D)|$ 153

D.17 Comparison of data/simulation of $p_t(D)$ and $\min p_t$ 153

D.18 Comparison of data/simulation of min $|d_0|$ 154

D.19 Comparison of data/simulation of correlations, part1 156

D.20 Comparison of data/simulation of correlations, part2 157

D.21 Comparison of data/simulation of correlations, part3 158

D.22 Comparison of data/simulation of correlations, part4 159

D.23 Comparison of data/simulation of $|d_0(B)|$ and $|d_0(D)|$ 160

D.24 Comparison of data/simulation of $\chi^2_{3D}(B)$ and $\chi^2_{r\phi}(B)$ 161

D.25 Comparison of data/simulation of $L_{xy}/\sigma_{L_{xy}}(B)$ and $L_{xy}/\sigma_{L_{xy}}(D)$ 161

D.26 Comparison of data/simulation of $L_{xy}(B \to D)$ and $p_t(\pi_B)$ 162

D.27 Comparison of data/simulation of min $\frac{|d_0|}{\sigma_{d0}}$ and min p_t 162

E.1 Template fit for combinatorial background in lifetime space 164

E.2 Fit results for combinatorial background in lifetime space 165

E.3 Lifetime templates for misassigned final states 165

List of Tables

1.1 Summary of measurements that constrain the unitarity triangle 26

2.1 Accelerator parameters for the Tevatron. 29

4.1 Preselection cuts 56

6.1 Correlation of all training variables to the training target 89

6.2 Correlation of the training variables to the training target 92

6.3 Loss of total correlation for the sideband training 106

6.4 Training variables used for $B_s \to D_s \pi$ 110

7.1 Summary of fit results 124

B.1 Example for a variable significance table. 135

Introduction

The universe that surrounds us is composed of elementary particles known as quarks and leptons. These particles and the compounds they form, for example protons, atoms, stones or humans, are generally referred to as *matter*, mostly consisting of light quarks and electrons. But this is not the full story. It was suggested by Dirac in 1928 and then actually confirmed by Anderson in 1933 [16] that not only electrons but also anti–electrons, also called positrons, exist. The positron has the same properties as the electron, except for a charge of +1 and the fact that it's not made of matter but *antimatter*, which means that electron and positron will annihilate into two massless photons when they meet. It shows out that not only the electron has an antimatter partner, but that for all quarks and lepton there exist corresponding antiquarks and antileptons.
Quarks don't exist as free particles, they form bound states. Examples for such bound states are the proton, containing two u and one d quark, or the neutron, composed of one u and two d quarks. Generally bound states composed of three quarks are called *baryons*, bound states containing a quark and an antiquark form *mesons*.
Certain types of mesons show a particularly interesting behavior – they *oscillate*, which means that a meson can transform into its own antiparticle, the corresponding antimeson, by so–called flavor changing weak interactions. Within the Standard Model of particle physics this is explained by a difference between flavor and mass eigenstates, leading to a mass difference Δm between the meson's mass eigenstates. A *time–resolved* measurement of meson–antimeson oscillations can be used to determine Δm, which is directly proportional to the oscillation frequency. Meson–antimeson oscillations were first observed in 1956 in the $K^0 - \bar{K}^0$ system [7], which is composed of $\bar{s}d - s\bar{d}$ quarks. Other neutral mesons that have the potential to oscillate are the $D^0 - \bar{D}^0$ $(c\bar{u} - \bar{c}u)$, $B^0 - \bar{B}^0$ $(\bar{b}d - b\bar{d})$ and $B_s - \bar{B}_s$ $(\bar{b}s - b\bar{s})$ systems. While in the case of the $D^0 - \bar{D}^0$ system the mass difference between both states is predicted to be very small, leading to a very small oscillation frequency which renders the oscillations difficult to observe, the oscillations of the B^0 and B_s mesons are long established. The first evidence of neutral B meson oscillations were observed in 1987 by the UA1 [8] and ARGUS [9] experiments. Today very precise time–resolved measurements of $B^0 - \bar{B}^0$ oscillations exist and provide a world average value of $\Delta m_d = 0.507 \pm 0.004$ ps^{-1} [10].

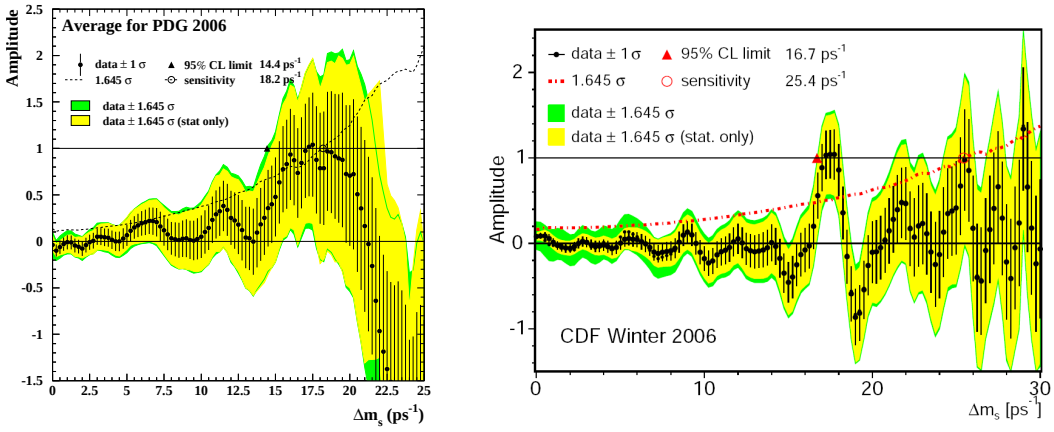


Figure 6: Amplitude scan for Δm_s using all available experimental results prior to 2006 (left) [10] and the amplitude scan from CDF II (right) [11].

A time–resolved measurement in the $B_s - \bar{B}_s$ system proves to be more difficult due to the much more rapid oscillations and the smaller production rate compared to the B^0 system. Until very recently only an upper limit of $\Delta m_s > 14.4$ ps^{-1} [10] on the mixing frequency Δm_s was known, based on the results of LEP, SLD and CDF. At the time of this write–up, for the first time a direct measurement of time–resolved B_s could be realized by the CDF II collaboration, quoting a value of $\Delta m_s = 17.33^{+0.42}_{-0.21} \pm 0.07$ ps^{-1} [11]. The significance of the B_s oscillation signal is 3.5 σ and the probability of a background fluctuation producing this kind of signal is 0.5 %. Figure 6 compares the amplitude scans used to derive the limit of 14.4 ps^{-1} (left) to the amplitude scan obtained as part the CDF II result (right). The CDF II Collaboration (CDF means "Collider Detector experiment at Fermilab") is an international collaboration of about 500 Physicists and 60 universities and national laboratories. The aim of the work presented in this thesis is to further improve the measurement of Δm_s at CDF II by introducing a neural network algorithm for an improved candidate selection with the aim to separate signal from background events more efficiently. The Δm_s signal significance scales with $\frac{S}{\sqrt{S+B}}$, where S denotes the number of signal events observed in a data sample on a background level of B events. Using a neural network to improve the signal significance $\frac{S}{\sqrt{S+B}}$ can lead to a more significant measurement of Δm_s and help to exclude the possibility of a background fluctuation.
In addition to an improved signal selection, also an unbinned maximum likelihood fit for a direct measurement of Δm_s is implemented as part of the work presented in this thesis. The fit is used to demonstrate that a neural network selection can be easily integrated into a measurement of Δm_s. In addition, the unbinned maximum likeli-

hood fit framework can be used for an independent cross check of the results presented in [11].
Both improved signal selection and mixing fit framework can be used in many applications beyond the scope of a B_s mixing measurement. Assuming Δm_s to be known, its value can be fixed and the fit framework can be used to calibrate advanced tagging algorithms on B_s samples. These taggers can be applied in measurements of the CP asymmetry in the B_s system. The improved signal selection achieved by neural network algorithms can be used for any kind of analysis involving the decay of B_s mesons into exclusive final states. Examples for such applications could be the measurement of branching fractions or spectroscopy in the B_s system.
The work presented in this thesis is mainly focused on the application in a Δm_s measurement. Chapter 1 starts with a general theoretical introduction on the unitarity triangle with a focus on the impact of a Δm_s measurement. Chapter 2 then describes the experimental setup, consisting of the Tevatron collider and the CDF II detector, that was used to collect the data. In chapter 3 the concept of parameter estimation using binned and unbinned maximum likelihood fits is laid out. In addition an introduction to the NeuroBayes® neural network package is given. Chapter 4 outlines the analysis steps walking the path from the trigger level selection to fully reconstructed B mesons candidates. In chapter 5 the concepts and formulas that form the ingredients to an unbinned maximum likelihood fit of Δm_s (Δm_d) from a sample of reconstructed B mesons are discussed. Chapter 6 then introduces the novel method of using neural networks to achieve an improved signal selection. First the method is developed, tested and validated using the decay $B^0 \to D\pi, D \to K\pi\pi$ and then applied to the kinematically very similar decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$. Chapter 7 uses events selected by the neural network selection as input to an unbinned maximum likelihood fit and extracts the B^0 lifetime and Δm_d. In addition, an amplitude scan and an unbinned maximum likelihood fit of Δm_s is performed, applying the neural network selection developed for the decay channel $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$. Finally chapter 8 summarizes and gives an outlook.

Chapter 1

Theoretical Overview

1.1 Overview

In the Standard Model of particle physics [1, 2, 3, 4] the strong, electromagnetic and weak interactions arise due to the exchange of spin–one bosons between spin–1/2 fermions. The gauge symmetry group underlying the Standard Model is $SU_C(3) \times SU_L(2) \times U_Y(1)$. The gauge bosons associated with each symmetry are

$$\begin{aligned} SU_C(3) &\longrightarrow 8\, G_\mu^\alpha, \quad \alpha = 1, \ldots, 8 \\ SU_L(2) &\longrightarrow 3\, W_\mu^\alpha, \quad \alpha = 1, 2, 3 \\ U_Y(1) &\longrightarrow B_\mu \end{aligned}$$

The eight spin–one particles G_μ^α denote gluons, the exchange particles of the strong interaction. The four bosons W_μ^α and B_μ give rise to the four physical bosons $W^\pm$, Z^0 and γ that mediate the weak and electromagnetic interactions. The fermion sector can be classified into three families, consisting of six flavors of *leptons*, by definition leptons are spin–1/2 particles that don't participate in the strong interaction, and six flavors of *quarks*, that participate in the strong interaction. All fermions take part in the weak interaction. The coupling of the charged $W^\pm$ bosons to the fermions is the only interaction in the Standard Model that connects fermions of different families. Without charged current interaction, the lightest particle of any fermion family would be stable, as a consequence the exchange of $W^\pm$ bosons is responsible for a large fraction of observed particle decays.

$$\underbrace{\begin{pmatrix}\nu_e\\ e\end{pmatrix}\begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}\begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}}_{\text{Leptons}} \qquad \underbrace{\begin{pmatrix}u\\ d\end{pmatrix}\begin{pmatrix}c\\ s\end{pmatrix}\begin{pmatrix}t\\ b\end{pmatrix}}_{Quarks}$$

Figure 1.1: Three families of fermions grouped into three generations of leptons and quarks.

1.2 The CKM Matrix

The interaction Lagrangian describing the charged current interaction between different quark flavors can be written in terms of mass eigenstates as

$$\mathcal{L}_{cc} = -\frac{g_2}{\sqrt{2}}\left(\bar{u}_L, \bar{c}_L, \bar{t}_L\right)\gamma^\mu\left(1-\gamma^5\right)V_{CKM}\begin{pmatrix}d_L\\ s_L\\ b_L\end{pmatrix}W_\mu^\dagger + h.c.$$

where V_{CKM} is the 3×3 transformation matrix between quark mass and flavor eigenstates, called Cabibbo–Kobayashi–Maskawa (CKM) matrix [5, 6]. Most generally the CKM–matrix can be written as

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub}\\ V_{cd} & V_{cs} & V_{cb}\\ V_{td} & V_{ts} & V_{tb}\end{pmatrix}$$

The V_{ij} are complex numbers, leading to 18 real parameters of the matrix. Assuming three generations of quarks allows to impose the unitarity requirement $V_{CKM}^\dagger V_{CKM} = \mathbf{1}$, which eliminates 9 free parameters. Furthermore, multiplications of the different quark fields by a phase alter V_{CKM} but leave the overall Standard Model Lagrangian unchanged — modulo an overall rotation of all quarks by a common phase, which again leaves V_{CKM} unchanged (see e.g. [17]). The freedom to redefine the quark fields removes another five degrees of freedom from the CKM–matrix. Four physical parameters remain free, which can be interpreted as three rotational angles plus a complex phase.

A popular and useful parametrization of the CKM–matrix was derived by Wolfenstein [18], which indicates the size of each matrix element by the small quantity $\lambda \simeq 0.2$ up to third order:

$$V_{CKM} \simeq \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta)\\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2\\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1\end{pmatrix} + \mathcal{O}(\lambda^4) \qquad (1.1)$$

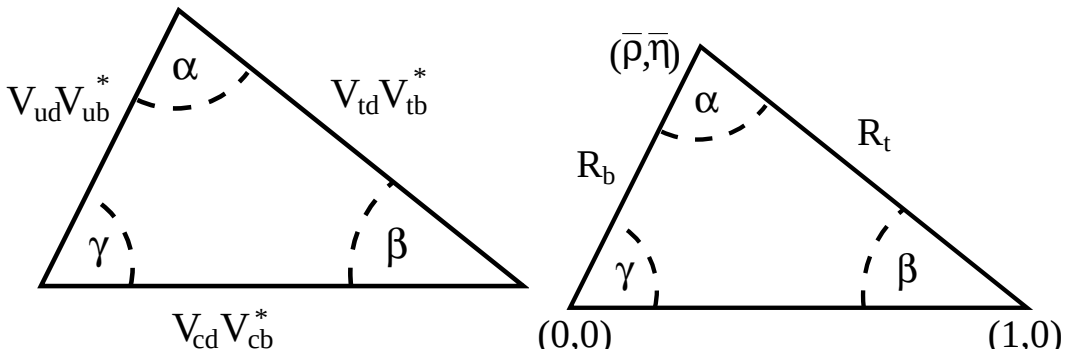


Figure 1.2: The unitarity triangle in standard form (left) and rescaled by $V_{cd}V_{cb}^*$ in the Wolfenstein representation (right).

1.3 The Unitarity Triangle

The matrix elements of V_{CKM} obey unitarity constraints, which means that any pair of rows or columns is orthogonal. From an experimental point of view, out of the six arising orthogonality conditions the most interesting is the product of the first and third column of V_{CKM}.

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$$

This condition can be displayed as a closed *unitary triangle* in the complex plane. All sides of this triangle are of the same order of magnitude and the angles α, β and γ relate to CP–violating quantities that can be measured experimentally. For a more convenient representation it is useful to scale all sides by $V_{cd}V_{cb}^*$. Moving to the Wolfenstein parametrization (eq. 1.1) the corners of triangle have the coordinates $(0,0)$, $(1,0)$ and $(\bar{\rho},\bar{\eta})$, where $\bar{\rho},\bar{\eta}$ are related to the Wolfenstein parameters as

$$\bar{\rho} = \left(1-\lambda^2\right)\,\rho, \quad \bar{\eta} = \left(1-\lambda^2\right)\,\eta$$

Figure 1.2 depicts the unitarity triangle in standard and in rescaled Wolfenstein representation.

1.4 Constraining the Unitarity Triangle

Measuring the sides and angles of the unitarity triangle has been one of the major tasks of experimental particle physics in recent years. Overconstraining the unitarity triangle by measuring the sides and angles with different, independent measurements will prove an important test for the Standard Model. If it is experimentally found that the triangle is not closed, this will be a hint for new physics beyond the Standard

Model.
Today, most experimental results that can be used to constrain the coordinates $(\bar{\rho}, \bar{\eta})$ of the upper edge of the unitarity triangle depend on hadronic matrix elements which introduce a significant uncertainty to the extracted value of the parameters. Figure 1.3 summarizes measurements of quantities that at present provide the most stringent constraints on the unitarity triangle: $|V_{ub}/V_{cb}|$, $\sin 2\beta$, ϵ_K and the ratio $\Delta m_s/\Delta m_d$.

This section summarizes the current experimental status of the measurements of $|V_{ub}/V_{cb}|$, $\sin 2\beta$ and ϵ_K before Δm_s and Δm_d are discussed in more detail, due to their relevance for the work presented in this thesis.

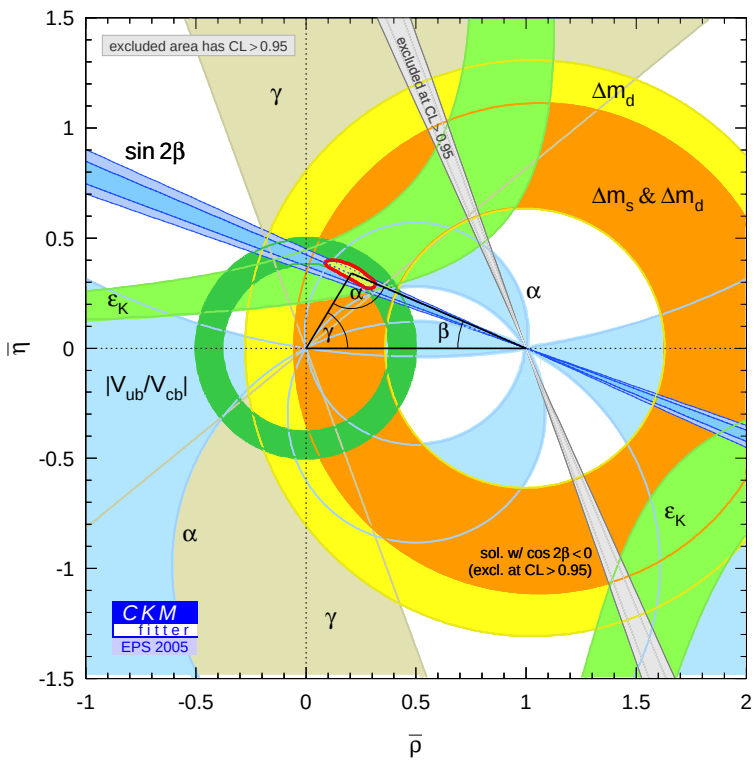


Figure 1.3: Graphical summary of the experimental constraints on the parameters of the unitarity triangle [19]. This does not yet include the recent Δm_s measurement performed by CDF II.

1.4.1 $|V_{ub}/V_{cb}|$

The parameter $|V_{ub}/V_{cb}|$ can be determined from the relative rate of the semileptonic b–quark decays $b \to cl\nu$ and $b \to ul\nu$. Currently the most precise measurements of V_{cb} result from the study of semi–exclusive decays $B \to D^{(*)}\, l\, \bar{\nu}$ and inclusive B to charm decays. Similarly, V_{ub} can be obtained from measuring the decays $B \to m\, l\, \bar{\nu}$, where m denotes a π, ρ, ω or η in the exclusive final state. The most precise measurements of these decays are currently provided by the BaBar and CLEO experiments [20][21]. The constraint on $|V_{ub}/V_{cb}|$

$$\frac{1}{\lambda}\left(1 - \frac{\lambda^2}{2}\right)\left|\frac{V_{ub}}{V_{cb}}\right| = \sqrt{\rho^2 + \eta^2} \tag{1.2}$$

describes a circle passing through $(\bar{\rho}, \bar{\eta})$ centered around $(0, 0)$, as seen in figure 1.3.

1.4.2 $\sin 2\beta$

The time dependent CP–violation parameter S, which can be measured in $b \to c\bar{c}s$–type decays with and without mixing, is equal to $\sin 2\beta$ to a very good precision [19]. The world average given by $\sin 2\beta = 0.739 \pm 0.048$ [22] is dominated by measurements performed by BaBar and Belle. As there are no uncertain hadronic matrix elements involved in the measurement of $\sin 2\beta$, this measurement provides a very stringent constraint on the position of $(\bar{\rho}, \bar{\eta})$ in the complex plane.

1.4.3 ϵ_K

The CP–violation parameter ϵ_K is a measure of CP–violation in the kaon sector. It can be extracted from measurements of the ratios of amplitudes η_{00} and η_{+-} of the neutral kaons K_L^0, K_S^0 decaying to a pair of neutral or charged pions respectively:

$$\epsilon_K = \frac{1}{3}\eta_{00} + \frac{2}{3}\eta_{+-}$$

The current PDG [15] values for η_{00} and η_{+-} lead to an average of $\epsilon_K = (2.282 \pm 0.017) \times 10^{-3}$ [19].

1.4.4 Δm_d and Δm_s

In analogy to the neutral kaon sector, two CP conjugate states exist in the neutral B meson system. The mass eigenstates are not eigenstates of the weak interaction, but mixtures of the two CP conjugate states. Mixing between these states emerges via the two box diagrams shown in figure 1.4.

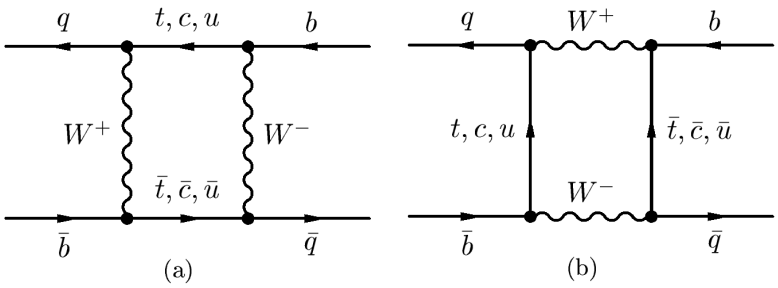


Figure 1.4: Diagrams giving rise to B mixing. The label q can represent either a d–quark for B^0 mixing or a s–quark in the case of the B_s.

The flavor eigenstates $\left|B_q^0\right\rangle$ and $\left|\bar{B}_q^0\right\rangle$ can be written in terms of the mass eigenstates $|B_{H,q}\rangle$ (heavy) and $|B_{L,q}\rangle$ (light) as

$$\left|B_q^0\right\rangle = \frac{1}{\sqrt{2}}\big(\,|B_{H,q}\rangle + |B_{L,q}\rangle\,\big)$$
$$\left|\bar{B}_q^0\right\rangle = \frac{1}{\sqrt{2}}\big(\,|B_{H,q}\rangle - |B_{L,q}\rangle\,\big)$$

Here q can represent either a d–quark for B^0– or an s–quark for B_s–mixing.
The time evolution of a state $|s\rangle$ in the non–relativistic approximation is described by the Schrödinger equation

$$H_{tot}\,|s\rangle = i\frac{\partial}{\partial t}\,|s\rangle$$

The evolution in time for an initially pure B_q^0 ($\bar{B}_q^0$) state is given by

$$\left|B_q^0(t)\right\rangle = e^{-\Gamma_q t/2}e^{-im_q t}\big(\cos(\Delta m_q\,t/2)\left|B_q^0\right\rangle + i\sin(\Delta m_q\,t/2)\left|\bar{B}_q^0\right\rangle\big)$$
$$\left|\bar{B}_q^0(t)\right\rangle = e^{-\Gamma_q t/2}e^{-im_q t}\big(i\sin(\Delta m_q\,t/2)\left|B_q^0\right\rangle + \cos(\Delta m_q\,t/2)\left|\bar{B}_q^0\right\rangle\big)$$

The difference of the widths Γ_d between short– and long–lived state of the B_s meson is expected to be small in the case of the B^0 meson [23] and can be neglected. Therefore the probability for an initially pure B^0 ($\bar{B}^0$) to decay as $\bar{B}^0$ (B^0) is

$$P(t) = \frac{1}{2}e^{-\Gamma_d t}\big(1 - \cos(\Delta m_q\,t)\big) \tag{1.3}$$

The width difference $\Delta\Gamma_s$ between short– and long–lived state of the B_s mesons is predicted to be significantly larger than $\Delta\Gamma_d$ [23], current experimental results give $\Delta\Gamma_s/\Gamma_s = 0.31^{+0.10}_{-0.11}$ [10]. Incorporating $\Delta\Gamma_s$ introduces an additional term in the

probability for an initially pure B_s ($\bar{B}_s$) to decay as $\bar{B}_s$ (B_s) [24]

$$P(t) = \frac{1}{2} e^{-\Gamma_s t} \left(\cosh\left(\frac{e^{-\Delta\Gamma_s t}}{2}\right) - \cos(\Delta m_s\, t) \right) \tag{1.4}$$

Within the Standard Model, the theoretical prediction for the values of Δm_q can be calculated at lowest order evaluating the mixing diagrams in fig. 1.4. The result given in [25] is

$$\Delta m_q = \frac{G_F^2}{6\pi^2}\, \eta_B\, m_{B_q}\, f_{B_q}^2\, B_q\, m_W^2\, S(x_t)\, \left|V_{tb}^* V_{tq}\right| \tag{1.5}$$

Here m_{B_q} is the B_q^0 meson mass, $f_{B_q}\sqrt{B_q}$ is the hadronic matrix element taken from Lattice QCD [26] and η_B is a perturbative QCD correction to the Inami–Lim function $S(x_t)$ [27]. Translating the factor $|V_{tb}^* V_{td}|$ into the Wolfenstein parametrization shows that a constraint on Δm_d describes approximately a circle around (1,0) in the $(\bar{\rho}, \bar{\eta})$–plane:

$$V_{tb}^* V_{td} = \lambda^6 A^2 [(1-\bar{\rho}) + \bar{\eta}^2] + \mathcal{O}(\lambda^{10}) \tag{1.6}$$

The largest theoretical uncertainties in the expression for Δm_q in eq. 1.5 arise from the calculation of the hadronic matrix elements $f_{B_q}\sqrt{B_q}$. When measuring the mass difference Δm_q with high precision, the error on $f_{B_q}\sqrt{B_q}$ will still introduce a large uncertainty of the matrix element V_{tq} extracted from eq. 1.5. Therefore measurements of Δm_d or Δm_s alone don't additionally constrain the unitarity triangle in the $(\bar{\rho}, \bar{\eta})$–plane.

The importance of a measurement of Δm_d and Δm_s arises from the fact that most theoretical uncertainties cancel out when instead of the expression for Δm_q from eq. 1.5, the ratio $\Delta m_s/\Delta m_d$ is taken into account:

$$\frac{\Delta m_s}{\Delta m_d} = \frac{m_{B_s}\, f_{B_s}^2\, \sqrt{B_s}}{m_{B^0}\, f_{B^0}^2\, \sqrt{B^0}} \left|\frac{V_{ts}}{V_{td}}\right|^2 \tag{1.7}$$

The ratio of the matrix elements $\xi = f_{B_s}^2\sqrt{B_s}/f_{B^0}^2\sqrt{B^0}$ can be calculated from lattice QCD with an uncertainty below 5 % [28]. Measuring the mass differences Δm_s and Δm_d allows to extract the ratio of the CKM matrix elements V_{ts}/V_{td}.

Very recently, the first direct measurement of Δm_s was published by the CDF II collaboration [11]. The result $\Delta m_s = (17.33^{+0.42}_{-0.21} \pm 0.07)\,\mathrm{ps}^{-1}$ can be combined with the current world average values of $\Delta m_d = (0.507 \pm 0.004)\,\mathrm{ps}^{-1}$, $m(B^0) = 5.2794 \pm 0.0005\,\mathrm{GeV}$ [1], $m(B_s) = 5.3696 \pm 0.0024\,\mathrm{GeV}$ [10] and $\xi = 1.21^{+0.040}_{-0.024}$ [28] to obtain $\left|V_{ts}/V_{td}\right| = 0.2087^{+0.0047}_{-0.0076}$.

[1] In this work energy, momentum and mass are given in the unit GeV. One GeV corresponds to $1.602\,176\,462(63) \cdot 10^{-10}$ Joule, obtaining momentum (mass) in the unit $\frac{kg\,m}{s}$ (kg) is realised by the division by c (c^2), where $c = 299\,792\,458\,\frac{m}{s}$.

The measurement of Δm_s from [11] is already very precise and the value derived for $\left|V_{ts}/V_{td}\right|$ is now already limited by the theoretical precision of ξ. Providing a very precise result for Δm_s on the one hand, the measurement published by CDF II observes a signal with a significance of 3.5 σ only on the other hand. As already mentioned in the introduction, the probability of a background fluctuation producing a spurious signal is still 0.5 % and can therefore not be completely excluded.
The main focus of the work presented in this thesis is to improve the significance of the existing Δm_s measurement by selection signal events more efficiently and at the same time suppress background contributions. The selection algorithm suggested in this thesis will therefore help to exclude the possibility of a background fluctuation and will provide an important contribution to fully establish a signal for time resolved B_s oscillations in the near future.

Parameter	Value
V_{ub}	$(3.90 \pm 0.68) \times 10^{-3}$
V_{cb}	$(42.0 \pm 0.8) \times 10^{-3}$
$\lvert\epsilon_K\rvert$	$(2.282 \pm 0.017) \times 10^{-3}$
$\sin 2\beta$	0.739 ± 0.048
Δm_d	$(0.507 \pm 0.004)\,\mathrm{ps}^{-1}$
Δm_s	$(17.33^{+0.42}_{-0.21} \pm 0.07)\,\mathrm{ps}^{-1}$

Table 1.1: Summary of measurements of some important parameters that constrain the unitarity triangle [10] [11] [19].

Chapter 2

Experimental Setup

2.1 The Tevatron

The Tevatron proton–antiproton ($p\bar{p}$) collider is located at the Fermi National Accelerator Laboratory in Batavia/Illinois. The main accelerator ring has a radius of 1 km and is operated at a center of mass energy of 1.96 TeV, which makes it the source of the world's highest energy $p\bar{p}$–collisions today. The collisions occur at two interaction regions where the two detectors D0 and CDF II ("Collider Detector at Fermilab") are located. For this thesis data collected with the CDF II detector is used.

2.1.1 The Accelerator Chain

To accelerate protons and antiprotons to a center of mass energy of 1.96 TeV, a chain of successive preaccelerators (figure 2.1) is used before finally injecting the particles into the Tevatron. As a first stage hydrogen gas is ionized to H^- and then accelerated to a kinetic energy of 750 keV inside the *Cockroft–Walton* device. From there the ions enter a linear accelerator (*Linac*) which uses oscillating electric fields in RF cavities to accelerate the ions traversing a distance of 150 m to 400 MeV. Due to the geometry of the oscillating RF fields, the ions are grouped into bunches at the end of the Linac. At this stage, the ions are directed onto a carbon foil target which strips the hydrogen off its electrons, leaving bare protons.
Next, the *Booster*, a circular synchrotron 74.5 m in diameter, accelerates the protons by a series of kicks from RF cavities to 8 GeV, with each turn around the booster the protons gain 500 keV of kinetic energy. Here the intensity of the proton beam is increased by repeatedly injecting further protons into the same orbit as the already circulating ones. Protons are then extracted from the Booster into the *Main Injector*,

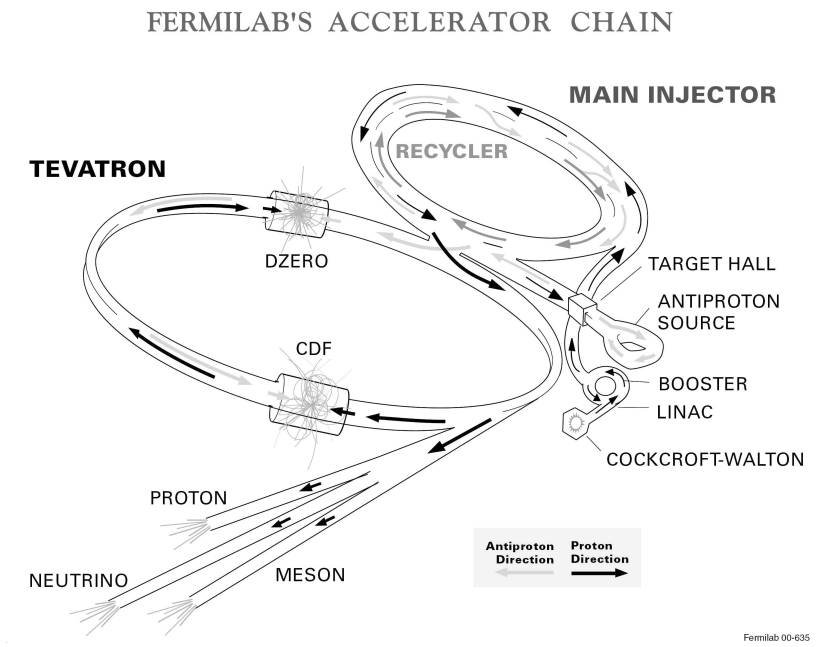


Figure 2.1: The Fermilab accelerator complex for Run 2.

a 3 km circumference synchrotron, where protons can be accelerated to 120 GeV for antiproton production or up to 150 GeV for injection into the Tevatron.
Antiprotons are produced by directing the 120 GeV protons from the Main Injector onto a nickel target, where ~ 20 antiprotons with a kinetic energy of 8 GeV are produced per 10^6 protons. The antiprotons are then separated from other particle species by a pulsed magnet and focused by a lithium lens, and via the Debuncher Synchrotron accumulated in the 8 GeV *Accumulator Synchrotron*, where stochastic cooling is used to reduce the spread in the antiprotons kinetic energy spectrum before they can be directed to the Main Injector.
The production rate of antiprotons is the limiting factor for operating at high luminosities, it takes about 15–20 hours to build up a stack of antiprotons that will be injected into the Tevatron. Approximately once per day the 36 bunches of stacked antiprotons ($\sim 5 \times 10^{10}$ particles) and protons ($\sim 3 \times 10^{11}$ particles) are injected from the Main Injector into the Tevatron where they are accelerated to the final center of mass energy of 1.96 TeV. After the end of a collision period, also called *store*, the remaining unused antiprotons are collected and cooled in the *Antiproton Recycler*, which is housed in Main Injector ring and stores the particles for *recycling* them in the next store.

2.1.2 Performance

Since Winter 2004 the Antiproton Recycler is fully commissioned and integrated into the Tevatron operation. As $\sim$ 75 % of all antiprotons are expected to survive a store, and the production rate of antiprotons is the limiting factor to achieve high luminosities, the recycling of antiprotons resulted in a significant improvement of the performance in terms of peak luminosity and the Run 2 design luminosity of $\mathcal{L} = 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ could be reached. The luminosity of collisions can be written as:

$$\mathcal{L} = f\,F\left(\frac{\sigma_L}{\beta^*}\right)\frac{N_B N_p N_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} \tag{2.1}$$

where f is the revolution frequency, F is a form factor describing the geometric properties of a bunch, N_B the number of bunches, $N_{p,\bar{p}}$ the number of protons/antiprotons per bunch and $\sigma_{p,\bar{p}}$ the RMS beam size at the interaction point.
The *integrated luminosity* over the time of a data taking period multiplied with the cross section σ of a given process yields an estimate for the predicted number of events $N = \sigma \times \int \mathcal{L}\,dt$.
Table 2.1 summarize the values for those parameters for Run 2. Figures 2.2 and 2.3 summarize the development of peak and integrated luminosity at the CDF II detector since the beginning of Run 2 in 2002.

Parameter	Value for Run 2
N_B	36
bunch length [m]	0.37
bunch spacing [ns]	396
protons/bunch (N_p)	$\sim 3 \times 10^{11}$
antiprotons/bunch ($N_{\bar{p}}$)	$\sim 5 \times 10^{10}$
highest peak $\mathcal{L}$ [$\mathrm{cm}^{-2}\,s^{-1}$]	1.72×10^{32}

Table 2.1: Accelerator parameters for the Tevatron.

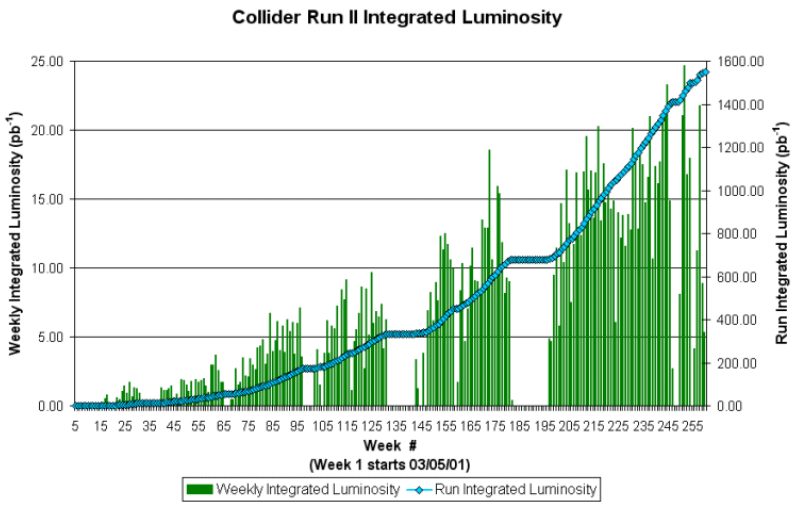


Figure 2.2: Integrated Tevatron Run 2 luminosity.

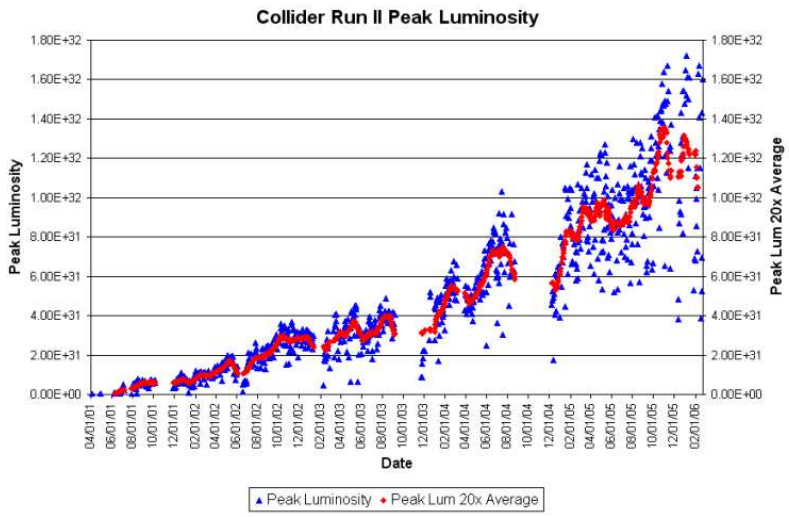


Figure 2.3: Tevatron Run 2 peak luminosity.

2.2 The CDF II Detector

2.2.1 Overview

The CDF II detector is a multipurpose collider detector [29], designed to detect and measure properties of particles being produced in $p\bar{p}$–collisions. It features a vertexing and tracking system, particle identification, a superconducting solenoid generating a 1.4 T magnetic field, calorimetry and muon chambers. The components are arranged in the cylindrical symmetry typical to collider detectors. Fig. 2.4 shows the side view one quadrant of the CDF II detector.

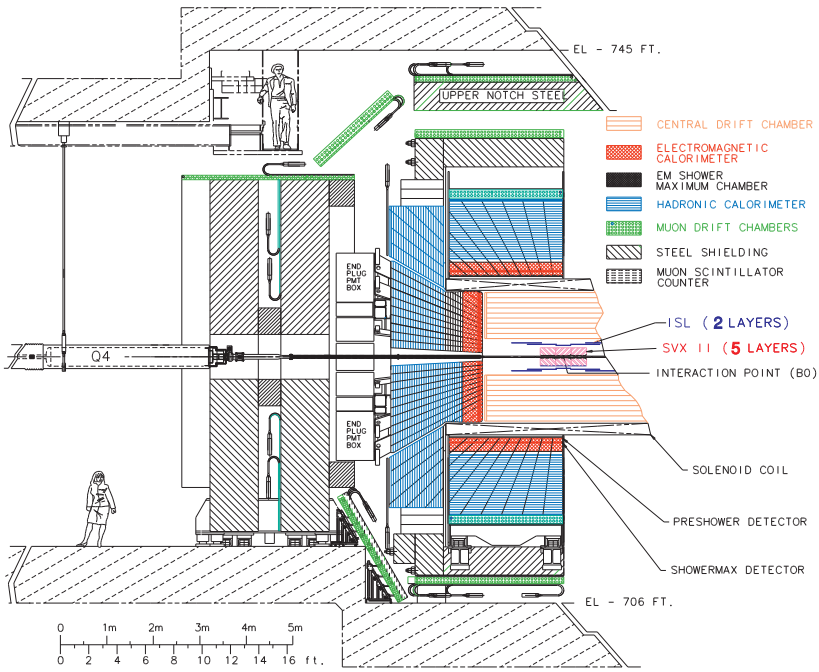


Figure 2.4: The CDF II detector.

2.2.2 The Tracking System

The CDF II tracking system consists of several subdetectors that can be used to detect charged particles and measure momenta and displacements with respect to the collision point (primary vertex) in the detector. It consists of two detectors, the *Central Outer Tracker* (COT) and the *Silicon Vertex Detector* with its subsystems SVXII, ISL (*Intermediate Silicon Layer*) and L00 (*Layer00*). The tracking system is built cylindrically around the beam pipe and contained within the 1.4 T magnetic field created by a superconducting solenoid.

The Silicon Vertex Detector of CDF II covers a radial range from r=1.6 cm to 28 cm, provides an acceptance up to $|\eta| < 2.0$ and consists of 6–7 layers of double and single layered silicon sensors. The innermost single layer closest to the beam line at a radius of 1.6 cm is called Layer00. Outside of L00 from r=2.1 cm to r=17.3 cm follows SVXII, a double-sided silicon microstrip detector. The strips are aligned axially on one side, with 90-degree stereo on the other side for layers 0, 1, and 3 and small-angle stereo (1.2 degrees) on the other side for layers 2 and 4. Between SVXII and the Central Outer Tracker (COT), the *Intermediate Silicon Layer* enhances linking of tracks measured in both detectors. The ISL central layer is located at r=22 cm and forward/backward layers are at r=20 cm and r=28 cm. The central layer covers $|\eta| < 1.0$ and the forward/backward layers cover $1.0 < |\eta| < 2.0$. The ISL is a double-sided silicon microstrip detector (axial on one side, small-angle stereo on the other side).
The ensemble of silicon vertex detector devices is particularly suited to measure the impact parameter d_0 and the azimuthal angle ϕ of tracks with high precision and can therefore determine the decay vertex of long lived unstable particles with high spatial resolution.

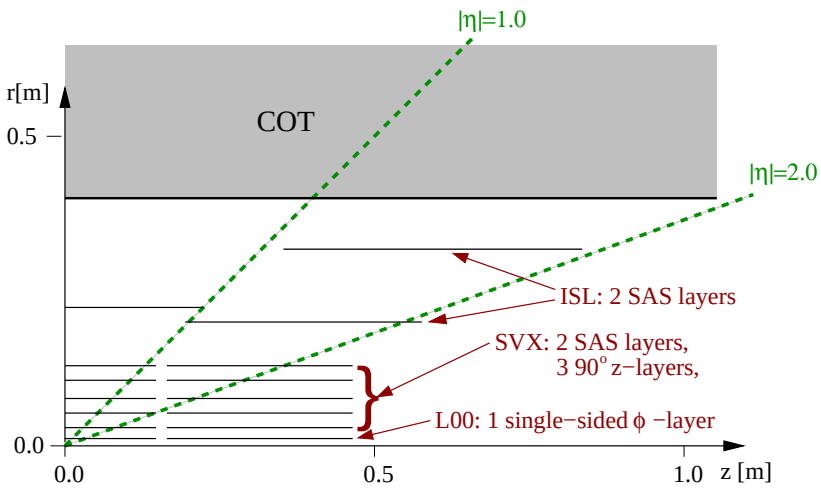


Figure 2.5: η coverage of the SVX and COT detectors.

The *Central Outer Tracker* (COT) is a cylindrical drift chamber covering the radial range from 33 cm up to 143 cm. The COT contains 96 wire layers grouped into eight superlayers, where in four superlayers the wires run in z–direction ("axial superlayers") and in the other four superlayers the wires are tilted by 2 degree with respect to the z–direction ("stereo superlayers"). All sense wire are contained in an Ar–Ethane–CF_4 gas mixture. Particles passing through the COT in the range $|\eta| < 1$ pass through all

eight superlayers, particles up to $|\eta| < 1.3$ pass through at least four superlayers and ionize the gas mixture. Electrons drifting towards the sense wire induce an electronic signal that is read out by an ASDQ (amplifier, shaper, discriminator and charge encoding) chip [30]. Due to geometrical reasons the track density in the COT tracking volume is significantly smaller compared to the SVX, in addition the COT produces a maximum of 96 hit measurements per tracks, so it provides an accurate measurement of a particle's transversal momentum (in the r–ϕ plane) and less precise information of the particle's momentum in the r–z plane.

2.2.3 Particle Identification

The ability to identify a track as coming from a kaon, proton, electron, muon or pion can yield an important input to many types of physics analyses. The CDF II detectors provides two measurements that can help to make a decision for each track regarding the type of particle that produced it, the measurement of the specific ionization (*COT dE/dx*) of charged particles inside the COT and the measurement of the particle's time of flight using the Time–of–Flight–detector (*TOF*) [31].
The specific ionization of charged particles transversing the COT gas mixture it measured by the readout chips of the sense wires. The pulse width read out is logarithmically dependent to the total charge deposited on the sense wire ($\Delta t \sim log(Q) \sim dE/dx$) [29]. To measure the time of flight T of a particle, information of the production time t_0 is necessary. For each collision event a measured TOF signal has to be matched to a corresponding COT track. A matched track can then be assigned to a primary vertex, knowledge of the spatial position of the primary vertex and the reconstruction of the tracks momentum p can be used to calculate t_0 (for details on track reconstruction and momentum measurement refer to section 4.2.1). Via the relation

$$T = T_{meas} - t_0 = \frac{s}{c}\sqrt{1 + \frac{m^2}{p^2}} \tag{2.2}$$

the mass m, and therefore the identity, of the particle producing track and TOF signal can be calculated. Here s denotes the path length of the track, calculated from the distance between primary vertex location and position of the measured TOF signal in the scintillator bars and c the velocity of light. Figure 2.6 shows the position of the TOF detector outside the COT (left) and the separation power of the TOF measurement for kaon/pion (K/π), proton/kaon (p/K) and proton/pion (p/π) separation compared to the particle separation power achieved by COT dE/dx alone (right).

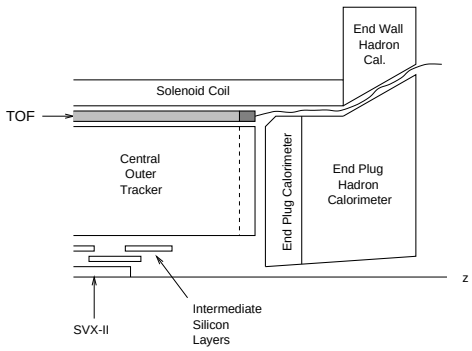


Figure 2.6: Location of the TOF system in the CDF II detector.

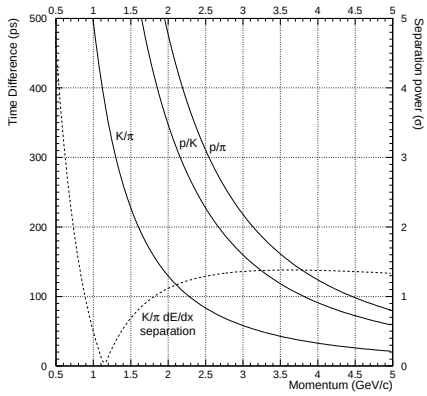


Figure 2.7: Time of flight differences as a function of particle type and momentum. The dashed line indicates K/π separation power of COT dE/dx.

2.2.4 Calorimeters

The CDF II calorimetry system consists of several independent calorimeters: the central and end–plug electromagnetic and hadronic calorimeters and the hadronic end–wall calorimeters. For the location of the individual detector components see figure 2.4. The calorimetry system covers the pseudorapidity region $|\eta| < 3.6$ and is used to measure the energy of high energy electrons, photons and hadronic jets.

2.2.5 Muon Systems

The CDF II muon system consists of four detectors mounted radially outside the calorimetry system. There are three drift chambers, namely the central muon detector, central muon upgrade and intermediate muon detector, and in addition a combined drift chamber/scintillation counter, called the central muon extension detector. Most of the particles that reach the muon system, i.e. without producing a shower and being absorbed in the calorimeters, actually are muons, as they interact with matter mainly via ionization. Kaons and pions surviving the calorimeter passage produce muon fake rates at the level of a few percent.

2.3 The CDF II Trigger System

2.3.1 Overview

Collisions at the CDF II detector occur at a rate of 2.5 MHz, the size of the detector information of one event is $\simeq$250 kB. Assuming that an interaction occurs at every collision, this would result in a data flow of 625 GB/s — the impossibility of storing every single event recorded by the CDF II detector on mass storage media is evident. Most events are minimum bias events that are not particularly interesting from a physical point of view. These events occur 4 orders of magnitude more frequently than hard inelastic processes producing b–quarks and even 10 orders of magnitude more frequently than processes producing top quarks. Therefore it is of utmost importance to do a preselection of events that are potentially interesting, which is performed by the CDF II trigger system, or short the *trigger*.
The trigger consists of three subsystems (*Levels*) that decide consecutively, the decisions being based on trigger tables that implement a large number of trigger paths, each path consisting of a set of rules that define requirements for events that can potentially be relevant for one of the different physics programs at CDF II. It is far beyond the scope of this short introduction to discuss all trigger paths used at CDF II, only a general introduction to the CDF II trigger system is given here. In a later chapter a focus will be put on the implementation of one trigger path that is of particular importance to the work presented in this thesis, the so–called *Two Track Trigger* (see section 4.1.1).

2.3.2 Level 1 Trigger

After each beam crossing, data from all detector components is read out, except for the data from the silicon vertex detector which is just sampled and kept in a pipeline, and made available for the decision of the Level 1 trigger. The core component of the Level 1 trigger system is the *eXtremely Fast Tracker (XFT)*. Based on a lookup table of pre–defined COT hit patterns, the XFT is capable of reconstructing the transversal momentum p_T and azimuthal angle ϕ of a track. At Level 1 also a matching of XFT tracks to muon chamber hits and calorimeter clusters is performed. As Level 1 takes a bit longer than a beam crossing to make its decision, the information from the different detector subsystems is kept in a 5 μs deep pipeline (see figure 2.8).

2.3.3 Level 2 Trigger

After an event has been accepted by Level 1, information from the silicon vertex detector (SVX) is digitized, read out and stored in one of the four Level 2 buffers. Those buffers allow Level 2 to make its decision asynchronously from Level 1, but can also

cause deadtime when all four buffers are full already and Level 1 accepts another event.
At Level 2 the *Silicon Vertex Tracker (SVT)* [32] takes the rather coarse information of the Level 1 XFT tracks and uses them as seed. Using sophisticated pattern–matching algorithms, the SVT tries to add silicon hit clusters to extrapolated XFT seed tracks. Using a linearized fit, the SVT creates tracks with a measurement of the impact parameter d_0 and significantly more precise values for p_T and ϕ. As shown in figure 2.11 the impact parameter resolution for SVT tracks is 47 μm. This resolution is a combination of the intrinsic SVT d_0–resolution and the transverse spread of the interaction region, which can be roughly described by a circular shape in the transversal plane with a Gaussian distribution of $\sigma \sim 33\,\mu m$. Taking into account this spread of the interaction region, the intrinsic SVT impact parameter resolution of $\sim 35\,\mu m$ is close to the resolution obtained after the offline reconstruction. The main purpose of the SVT is to preselect events with a positive lifetime by requiring impact parameters of $d_0 > 100\,\mu m$ for at least two SVT tracks, this requirement effectively selects tracks coming from particle decays with a positive lifetime, as demonstrated in figure 2.10. This is of particular importance when studying the decays of long lived b–hadrons.

2.3.4 Level 3 Trigger

The Level 3 trigger stage is implemented in software running on a PC farm. For the Level 3 trigger decision, the full COT and SVX offline reconstruction is performed for each track. Further improving the resolution of p_T, ϕ, z_0 and $\cot\theta$ and d_0, all Level 1 and Level 2 trigger conditions are confirmed with the resolution of offline tracks. If Level 3 accepts an event, the data is first stored on disk in a storage area and later written to tape for analysis. Roughly 1 in 120,000 events passes one or more paths of the full trigger chain, reducing the event output rate from the beam crossing rate to $\sim$ 100 Hz.

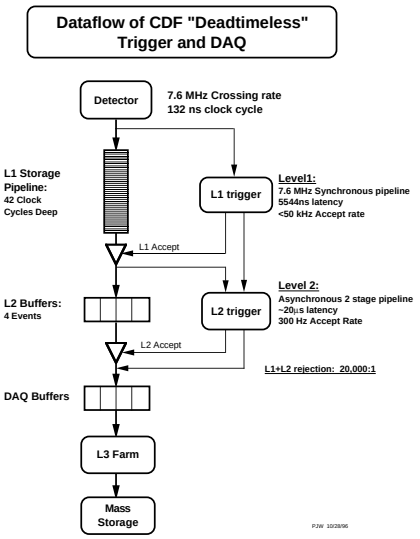


Figure 2.8: Data-flow in the CDF II trigger and data acquisition system.

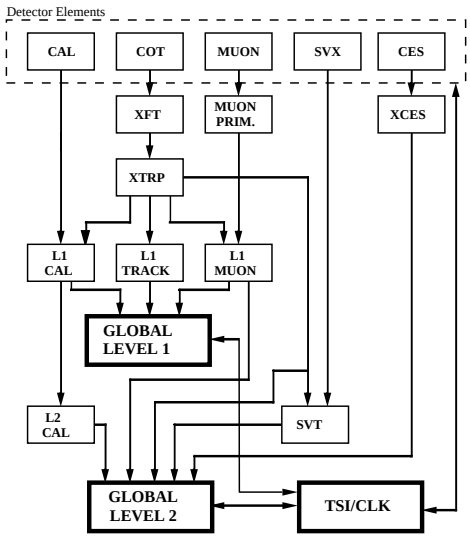


Figure 2.9: The first two trigger levels and the involved detector elements.

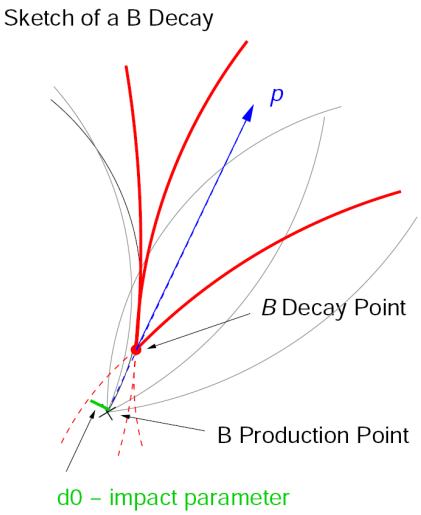


Figure 2.10: Definition of the track impact parameter.

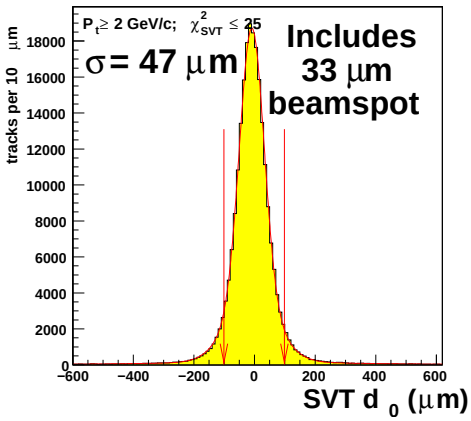


Figure 2.11: SVT impact parameter resolution.

Chapter 3

Mathematical Methods

3.1 Parameter Estimation

One of the challenges in quantitative science is the estimation of parameters from measurements with errors. These arise due to limited experimental resolution of the measurement process and are statistical in nature.
Let an experiment provide n measurements $\vec{x}_1, \vec{x}_2, ..., \vec{x}_n$ of the (multidimensional) variable $\vec{x}$. The task is to extractthe best estimate of $\vec{x}$ as well as the corresponding errors and correlation coefficients from these measurements. If the measurements follow a probability density function $f(\vec{x}|\vec{a})$, which is a function of the parameters $\vec{a}$, this problem can be solved by the maximum likelihood method [33].

3.1.1 The Maximum Likelihood Method

The maximum likelihood method assumes a multidimensional probability density function $f(\vec{x}|\vec{a})$ in $\vec{x}$ that depends on a set of unknown parameters $\vec{a}$ and, for a set of n measurements $\vec{x}_i$, defines the likelihood function $L(\vec{a})$

$$L(\vec{a}) = f(\vec{x}_1|\vec{a}) \cdot f(\vec{x}_2|\vec{a}) \cdot ... \cdot f(\vec{x}_n|\vec{a}) = \prod_{i=1}^{n} f(\vec{x}_i|\vec{a}) \tag{3.1}$$

The best estimate for the parameters $\vec{a}$ is the choice $\widehat{\vec{a}}$ that maximizes $L(a)$, i.e. maximizes the probability to observe the measured values $\vec{x}_i$. The maximum of $L(\vec{a})$ can be found by requiring

$$\frac{\partial L(\vec{a})}{\partial a_k} = 0 \qquad \text{for all } k \tag{3.2}$$

In real life $L(\vec{a})$ often tends to very small values causing numerical instabilities when evaluated on a computer. Instead of the likelihood function $L(\vec{a})$, also the log likelihood function $l(\vec{a}) = \ln L(\vec{a})$ can be used. As the logarithm is a monotonous function, requiring

$$\frac{\partial l(\vec{a})}{\partial a_k} = 0 \qquad \text{for all } k \tag{3.3}$$

is equivalent to the condition quoted in eq. 3.3. For practical purposes[1] it's useful to define the negative log likelihood function

$$F(\vec{a}) \;=\; -l(\vec{a}) \;=\; -\sum_{i=1}^{n} \ln f(\vec{x}_i|\vec{a}) \tag{3.4}$$

that has to be minimized to get the optimal estimate of $\widehat{\vec{a}}$. During the minimization procedure one has to make sure that the normalization in the fit region from $\vec{x}_{min}$ up to $\vec{x}_{max}$

$$\int_{\vec{x}_{min}}^{\vec{x}_{max}} f(\vec{x}|\vec{a})\, d\vec{x} = 1 \qquad \text{for all } \vec{a} \tag{3.5}$$

is kept constant and independent of $\vec{a}$. To obtain a correct result, one has to check that $f(\vec{x}|\widehat{\vec{a}})$ indeed describes the distribution of the measurements in each dimension properly when inserting the best estimate $\widehat{\vec{a}}$ of the parameters $\vec{a}$.
In the limit of an infinite number of measurements $n \to \infty$, the likelihood function approaches a Gaussian distribution. $F(\vec{a})$ can then be expanded around its minimum, where $\vec{a} = \widehat{\vec{a}}$ and $\frac{\partial F}{\partial \vec{a}} = 0$.

$$F(a_1, a_2, ..., a_N) = F(\widehat{a}_1, \widehat{a}_2, ..., \widehat{a}_N) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 F}{\partial a_i \cdot \partial a_j}(a_i - \widehat{a}_i)(a_j - \widehat{a}_j) + ... \tag{3.6}$$

$$= F(\widehat{a}_1, \widehat{a}_2, ..., \widehat{a}_N) + \frac{1}{2}\sum_{i,j} G_{ij}(a_i - \widehat{a}_i)(a_j - \widehat{a}_j) + ... \tag{3.7}$$

where $\vec{a} = (a_1, a_2, ..., a_N)$. In the limit $n \to \infty$ the matrix G can be identified as the inverse of the covariance matrix V^{-1}, also called *Hesse matrix* H. In the non–asymptotic case, V^{-1} serves as a good approximation of G at the minimum of F. In the case of only one parameter $F(a)$, the values of $F(a)$ around the minimum $a = \widehat{a} \pm r \cdot \sigma$ are the given by

$$F(\widehat{a} \pm r \cdot \sigma) = F(\widehat{a}) + \frac{1}{2}r^2 \tag{3.8}$$

[1] Several program libraries for *minimizing* functions exist, therefore it's just a matter of convenience to use a *negative* log likelihood function.

where σ can be identified as $\sigma = \frac{d^2F}{da^2}\Big|_{\widehat{a}}$. If $F(a)$ has the shape of a parabola around the minimum, equation 3.8 can be used to calculate the standard deviation. The probability that is contained in the interval $\widehat{a} \pm r \cdot \sigma$ can be derived from the Gaussian distribution. In the case of more than one parameter, the function

$$F_{min}(a_i) = \min F(\vec{a}) \tag{3.9}$$

can be defined for a_i, which is the minimum of $F(\vec{a})$ with regard to all other parameters a_j, $j \neq i$. From this definition, the $r\sigma$ standard deviations can be defined via

$$F_{min}(\widehat{a}_i + r\sigma) = F(\widehat{\vec{a}}) + \frac{1}{2}r^2 \tag{3.10}$$

similar to the case of only one variable. Non–parabolic behavior at the minimum can be approximated by asymmetric errors

$$F_{min}(\widehat{a}_i + \sigma_r) = F(\widehat{\vec{a}}) + \frac{1}{2} \tag{3.11}$$

$$F_{min}(\widehat{a}_i - \sigma_l) = F(\widehat{\vec{a}}) + \frac{1}{2} \tag{3.12}$$

3.1.2 The Maximum Likelihood Method for Binned Data

If the total number of measurements n is large, it's possible to process the result in form of a histogram where the x–axis is divided in J intervals (*bins*). The probability density function $f(x|a)$ is then fitted to the data points in each of the J bins. The entries n_j in each bin are random variables following a Poisson distribution with expectation value μ_j

$$P(n_j|\mu_j) = \frac{\mu^{n_j} e^{\mu_j}}{n_j!} \tag{3.13}$$

μ_i can be obtained from integrating $f(x|a)$ over the width of one interval multiplied with the size of the sample n

$$\mu_j = \int_{binj} f(x|a)\, dx \;\simeq\; n \cdot f(x_c|a) \cdot \Delta x \tag{3.14}$$

where x_c is the value of x at the center of bin j and Δx the width of the bin. The negative log likelihood function becomes

$$F(a) = -\sum_{j=1}^{J} \ln\left(\frac{\mu^{n_j} e^{\mu_j}}{n_j!}\right) \;=\; -\sum_{j=1}^{J} n_j \ln \mu_j + \sum_{j=1}^{J} \mu_j + \sum_{j=1}^{J} \ln n_j! \tag{3.15}$$

The last term $\sum_{j=1}^{J} \ln n_j!$ can be ignored during the minimization as it is constant for a given set of data points. In the case $\mu \gg 1$ the Poisson distribution can be approximated by a Gaussian distribution with variance $\sigma_i^2 = \mu_i$. The negative log likelihood function transforms to

$$F(a) = -\sum_{j=1}^{J} \ln\Big(\frac{1}{\sqrt{2\pi}\sigma_j} \exp\big(-\frac{(n_j - \mu_j)^2}{2\sigma_j^2}\big)\Big) \tag{3.16}$$

$$= \frac{1}{2}\sum_{j=1}^{J} \frac{(n_j - \mu_j)^2}{\mu_j} + \text{const.} \tag{3.17}$$

Here $2F(a)$ follows a χ^2 distribution with k degrees of freedom where k is the number of intervals J minus the number of parameters that are determined in the fit [33]. The value of χ^2/k should be $\simeq 1$, this can be used to decide whether the description of the data points by $f(x|a)$ is acceptable or not.

3.2 Neural Networks

The NeuroBayes® [34] neural network software combines a three–layer feed forward neural network with a robust and sophisticated preprocessing, which is performed before the training variables are actually fed into the network. Each training variable corresponds to one node in the input layer. Training patterns are passed to the network by feeding them into the input layer. Each node in the input layer is connected to the nodes in the intermediate or *hidden* layer. The number of nodes in the hidden layer can be chosen freely[2]. For a binary classification problem the output layer consists of only one node which gives a continuous network output in the interval $[-1, +1]$. Figure 3.1 shows a schematic representation of such a three layer network architecture. An additional bias node with a constant value of 1 in the input layer can help to increase the network performance. The nodes of two consecutive layers are connected via weights. For each node j a biased weighted sum depending on the values provided by the output of the nodes x_j of the previous layer is calculated

$$a_j(\vec{x}) = \sum_i w_{ij} x_i + \mu_{0,j} \tag{3.18}$$

The term $\mu_{0,j}$ represents the threshold of the node j provided by the connections of the bias node with the nodes in the intermediate layer and w_{ij} denote the weights

[2]The choice is not completely arbitrary. One has to make sure that the network has enough freedom to learn all features (not too few nodes) and prevent it from learning everything by heart (not too many nodes).

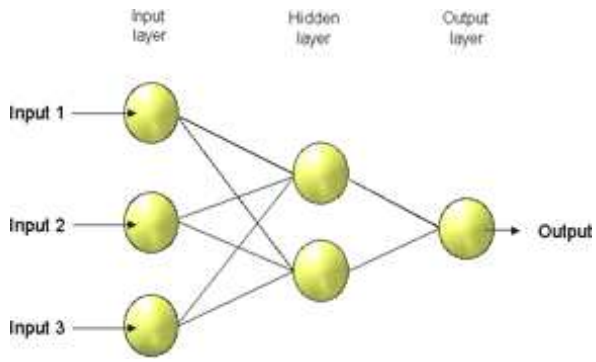


Figure 3.1: Schematic display of a three layer neural network architecture with one output node.

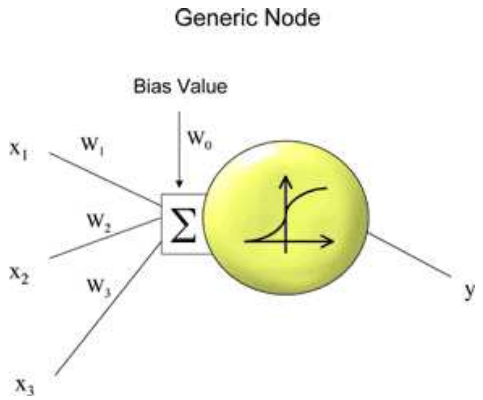


Figure 3.2: Representation of a neural network node. The weighted sum of the input variables x_i is passed through the sigmoid transfer function to the next layer.

from the input to the intermediate layer. The input to the next layer for each node is transformed by the symmetric sigmoid function

$$S(\vec{x}) = \frac{2}{1 + \exp(a(\vec{x}))} - 1 \tag{3.19}$$

The sigmoid transformation maps the interval $[-\infty, +\infty]$ to $[-1, +1]$, behaves linearly around 0 and goes into saturation for very small or large values of $a(\vec{x})$. For a network with d input nodes, M nodes in the hidden layer and vector of input values x_i, the network output of the k–th output node o_k is calculated via

$$o_k = S\Big(\sum_{j=1}^{M} v_{jk}\ S\big(\sum_{i=1}^{d} w_{ij}x_i + w_{0,j}\big)\Big) \tag{3.20}$$

where as before w_{ij} denotes the weights from the input to the intermediate layer and v_{jk} the weights from the intermediate layer to the output node.

3.2.1 Network Training

To train the neural network, training patterns are fed to the network via the input layer. For each training pattern the network output is calculated and compared to the true value T_{ij} for output node j and event i. Then the weights connecting different nodes in different layers are varied with the outline to minimize a cost function E. The cost function expresses the deviation of the actual network output from the desired or nominal value. In the work of this thesis, a cost function based on the entropy function is used in the network training:

$$E = \sum_i E_i = \sum_i \sum_j \ln\big(\frac{1}{2}(1 + T_{ij} o_{ij} + \epsilon)\big) \tag{3.21}$$

j sums over all patterns in the training sample, i over the output nodes and ϵ is a regularization constant that avoids numerical problems for an untrained network and is put to zero after a few iterations.
The surface of the cost function has a very complex shape in the multidimensional parameter space spanned by all possible combinations of weights. Finding the global minimum is practically impossible. At the beginning of the training, the weights are set to random values following a Gaussian distribution with mean 0 and width 1. The training is then performed by a special flavor of backpropagation and based on gradient descent, adjusting the change of each weight Δw_{ik} proportional to the current gradient of the cost function $\Delta w_{ik} = -\eta \frac{\partial E}{\partial w_{ik}}$. Hereby the step widths are adapted individually for each weight during the training.

3.2.2 Preprocessing of Variables

To aid the minimization of the cost function, NeuroBayes® performs a sophisticated preprocessing of the input variables. To avoid starting the training in the saturation region of the transfer function (eq. 3.19), the variables are first brought into the region between $[-1, +1]$ and then transformed into Gaussians with mean 0 and width 1. Then the variables are decorrelated by rotations of the covariance matrix, transforming it to a diagonal unit matrix which allows to minimize the weights independently. Furthermore the NeuroBayes® preprocessing is capable of treating variables that contain potentially ordered discrete classes and variables that don't have value assigned for each event by assigning the missing values to a δ function.
The significances of the input variables are determined during the preprocessing. The

correlation of each individual variable to the training target is calculated in an iterative procedure. One variable after the other is omitted from the set of training variables to determine the loss of total correlation caused by neglecting this particular variable. The variable causing the smallest information loss when omitted is then discarded, leading to a set of $n-1$ training variables. This procedure is repeated until the correlation to the target for variable is calculated. More details on the variable significance table can be found in appendix B.

3.2.3 Regularization

To enhance the generalization capabilities of the neural network and to avoid overtraining, several regularization schemes are applied during the network training. One way of improving the generalization ability is pruning away connections between nodes or even full nodes that have become insignificant during the network training process. The removal of connections and nodes reduces the number of free parameters and improves the ratio of signal over noise by removing the source of the noise.
Another regularization method helping to control the behavior of the network is weight decay. An additional weight is introduced to the error function from eq. 3.21

$$E \longrightarrow \widetilde{E} = E + \frac{c}{2}\sum_i w_i^2 \tag{3.22}$$

which effectively leads to a decay of weights during the minimization procedure, i.e. the network converges to smaller weights. This makes sure that only patterns occuring frequently are learned by the network, random fluctuations of the input data will decay away. In this way also very large weights are avoided that can increase the variance of the output and lead to discontinuities or oscillatory behavior of the network.

3.2.4 The Bayesian Approach

NeuroBayes® makes use of Bayes' theorem[35]

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \tag{3.23}$$

where H denotes a hypothesis and D the data that was measured. $P(H)$ includes a priori knowledge about the hypothesis, without knowing anything about the data, therefore also called *prior*. $P(D|H)$ is the likelihood to observe the D assuming H and $P(D)$ the probability to observe D. Finally $P(H|D)$ denotes the a posteriori knowledge about H that we have after taking into account the data, also called *posterior*. In the case that the training target is a real variable, NeuroBayes® can be used to estimate the full posterior probability density function $P(H|D)$ for each event. The

advantage of taking into account Bayes' theorem is that the output the neural network produces will always be in the range of the inclusive input distribution, and therefore will never produce non–physical results. In the case of binomial targets, the output of every well–trained neural network can be interpreted as Bayesian a posteriori probability in the case that the inclusive signal over background ratio is the same in the data sample that is classified and in the training sample.
Furthermore, Bayesian arguments are used in the automatic regularization and pruning schemes in NeuroBayes®.

Chapter 4

Dataset and Reconstruction

4.1 Dataset Selection

4.1.1 The Two Track Trigger

The data used for this work was collected entirely using the so–called *Two Track Trigger*, which is defined by a set of trigger paths based on the silicon vertex tracker (SVT, see section 2.3). The name Two Track Trigger indicates the dedication of this part of the trigger system: identify events with positive lifetime by selecting *two* displaced tracks (see figure 2.10). This type of trigger was first developed for the Run2 upgrade and is currently unique to CDF II. The Two Track Trigger requirements for an event can be summarized as:

- at least two oppositely charged XFT tracks with $p_t > 2\,\mathrm{GeV/c}$ and $\Delta\phi^{1,2} < 135°$ at Level 1.

- at least two SVT tracks with $p_t > 2\,\mathrm{GeV/c}$, $100\,\mu\mathrm{m} < |d_0| < 1\,\mathrm{mm}$ and $\chi^2_{\mathrm{SVT}} < 25$[1] at Level 2.

- match of the SVT tracks to COT tracks and confirmation of the p_t and impact parameter requirements with the additional requirements of $2° < \Delta\phi^{1,2} < 90°$ and the decay length in the r-ϕ-plane projected on p_t (L_{xy}) greater than 200 μm at Level 3.

[1]Here χ^2_{SVT} denotes the χ^2 of a track that is obtained after hits from the silicon detector are added to a Level 1 XFT track by the SVT, see section 2.3.3 for details.

Flexibly adapting the data taking bandwidth required by the Two Track Trigger to different luminosity scenarios, *additional* requirements are imposed on an event at the level 2 trigger stage by the individual Two Track Trigger subpaths:

L2_B_LOWPT requires at least two SVT tracks with $p_t^1 + p_t^2 > 4.0\,\text{GeV}$ (the so–called low–p_t scenario).

L2_B_CHARM asks for at least two oppositely charged SVT tracks with $p_t^1 + p_t^2 > 5.5\,\text{GeV}$.

L2_B_HIGHPT with at least two oppositely charged SVT tracks with $p_t^1 + p_t^2 > 6.5\,\text{GeV}$ (high–p_t scenario).

At periods of high luminosity data taking, the low–p_t scenario `B_LOWPT` [2] is usually *prescaled*. Prescaling means that only one in $N = 1/p$ events is accepted by the level 2 trigger systems out of all events passing the trigger requirements, p denoting the prescale factor. Applying a prescale factor to trigger paths at high luminosity data taking periods reduces the level 2 buffer occupancy, thereby reducing the overall deadtime of the trigger system. More bandwidth can then be given to trigger paths that select rare events at low rates. The prescale factors for individual trigger paths are adapted dynamically during a store[3] to take into account the decreasing luminosity by antiproton losses to make optimal use of the trigger system bandwidth. Also the `B_CHARM` and `B_HIGHPT` trigger paths may be subject to prescales. The prescale factors for all trigger paths during each data taking period are stored in a database and can be accessed in offline software. Figure 4.1 show the effective prescale factors (averaged per run) for the `B_LOWPT`, `B_CHARM` and `B_HIGHPT` trigger scenarios [36].

4.1.2 "Good Run" Selection

Figure 4.1 shows several run ranges where no prescale factors are quoted, this reflects the fact that not all runs should be used for any kind of analysis. Only data that is taken under stable conditions, where each trigger and detector component that is of importance for an analysis, was fully functional, should be used.
To keep track of problems with detector or trigger systems during individual data taking periods, several boolean good run flags are set by the shift crew after the completion of a run and stored in the online database.
The good run flags used for the work presented in this thesis are based on a set of requirements generally agreed on by the BPAK (B physics analysis kernel) group [37]

[2] The prefix `L2_` will be dropped from now on, the different Two Track Trigger scenarios don't differ at the L1 and L3 trigger stage.

[3] A store comprises one cycle of antiproton injections into the Tevatron, see section 2.1.

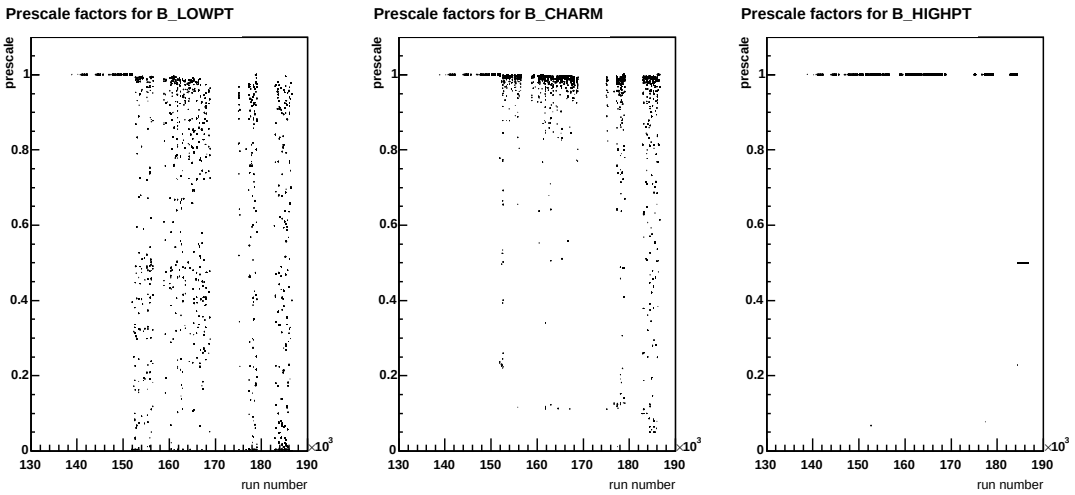


Figure 4.1: Effective prescale factors for the B_LOWPT, B_CHARM and B_HIGHPT trigger scenarios against run number. The run range displayed covers data taken from September 2002 until August 2004 is taken into account.

and reflect that for any analysis involving the reconstruction of B–mesons full functionality of the SVT trigger system and the tracking system is required. All flags applied for the dataset used in this work are summarized in the following list.

RUNCONTROL_STATUS During daily operation runs sometimes have to be ended very prematurely as problems with the trigger system, data acquisition or any detector system show up immediately after starting a run. Those runs can be easily identified by the very small number of events written out. The RUNCONTROL_STATUS flag is set to 0 if there are less than 10000 L1 accepts, less than 1000 L2 accepts and/or less than 1 nb^{-1} of data written to tape for this run.

SHIFTCREW_STATUS is set to 0 by the shift crew if a testing or calibration trigger table was used for this run. It is set to 1 for runs taken with approved physics trigger tables.

OFFLINE_STATUS is set by the offline shift crew, confirming that the data for a given run has passed offline processing without indicating any problems like corrupted data formats.

CLC_STATUS confirms the proper functionality of the Cerenkov luminosity counter *CLC* and that the luminosity measurement performed by the CLC was not affected by instable beam conditions.

L(1,2,3)T_STATUS assure proper operation of the trigger system. They are set by the consumer operator shift crew member after the comparison of reference quality monitoring plots for a given run to the default plots.

SVX_STATUS is set if the high voltage of the SVX silicon tracker is switched on.

SVT_STATUS is set based on quality monitoring plots of the corrected SVT beamline position and the online reconstruction of D mesons. Proper operation of the SVT requires the SVX to be switched on.

COT_OFFLINE sets a more rigorous limit on the minimum luminosity for a run of 10 nb^{-1} and requires the number of bad channels in the COT drift chamber to be less than 1 %.

Applying these good run requirements on the two datasets `xbhd0d` and `xbhd0h`, collected with the Two Track Trigger amounts to a collected integrated luminosity of 765 pb^{-1}. `xbhd0d` consists of data collected between February 2002 and August 2004 (355 pb^{-1}), `xbhd0h` includes data taken between December 2004 and September 2005 (410 pb^{-1}).

4.2 Event reconstruction

In the previous section the steps necessary to preselect a dataset of track pairs displaced from the primary interaction vertex, possibly originating from the decay of long–lived particles like B mesons, were discussed. Once these events are triggered and written to tape, they can be examined more carefully by using sophisticated reconstruction algorithms implemented in the CDF II offline software framework. In this section the algorithms involved in the proceeding from measuring hits in the tracking devices towards the reconstruction of B meson decays are discussed.

4.2.1 Tracking

Charged particles deposit small amounts of energy in the material of the CDF II tracking system. Due to the high voltages applied in the detector systems the resulting ionization of the detector material leads to small currents that can be detected and encoded by the detectors' readout systems. Such a signal obtained from one of the silicon sensors or COT readout wire is called a *hit*. The reconstruction of the trajectory of a charged particle by associating hits recorded in the detector systems to this particle is called track reconstruction or simply *tracking*. Several such tracking algorithms are used in CDF II offline software. As the track multiplicity is constant in ϕ, the hit occupancy reduces with $r\phi$ when a charged particle passes the detector

from the interaction point. The so–called *Outside–in (OI)* algorithms therefore start to reconstruct the flight paths of charged particles at the COT and then moves on by adding information from the SVX. Before the actual track fit is performed, two pattern recognition algorithms select track candidates in the COT. All candidates that pass the 2d–fit to a circle are then used for a 3d helix fit, making full use of the stereo layer information, and form a COT standalone track. The efficiency of COT track reconstruction is $\sim 95\,\%$ for track passing all eight superlayers with $p_t > 0.5$ GeV. The momentum resolution of $\sigma_{p_t}/p_t \simeq 0.15\,\%\,/\,\mathrm{GeV}$ is actually better than anticipated in the technical design report [29].
The next step to form an OI track is adding the information from the silicon tracking devices. The COT standalone track path is extrapolated until the intersection with layers of the silicon detector, if a silicon hit can be assigned to that track, the track fit repeated and the extrapolation to the next silicon layer is done. A track is considered an OI track if at least three $(r - \phi)$ silicon hits could be assigned to the COT standalone track. Two algorithms perform this work in the CDF II offline software framework, they are documented in detail in [38][13].

4.2.2 Momentum Scale Calibration and Track Refitting

As seen in the previous section, the energy loss by ionization of particles in matter forms the basis for reconstructing the flight paths in the detector. For a particle traversing the entire tracking volume, this energy loss is roughly 10 MeV. This energy loss however shows a dependence on the particle's momentum and rest mass[4], so in order to get an optimal momentum measurement, these effects have to be taken into account [39].
Figure 4.2 shows the p_t dependence of the reconstructed invariant mass of a J/ψ decaying into two muons. Two calibration tools can be used to remove the momentum dependence and to match the mass of the reconstructed J/ψ with the world average from [15]. First the momentum dependence is removed by modifying the incomplete detector description in the GEANT material map [40] used for track fitting. After removing the momentum dependence, the J/ψ mass is then shifted to the expected value by correcting the value used for the magnetic field B for the conversion of track curvature to momentum.
The average energy loss of a particle in the detector also depends in the particle's rest mass. Therefore, before reconstructing particle decays in dedicated final states, each track is refitted using the particle hypothesis expected in the reconstructed final state. Before the refit is performed, the COT covariance matrix is rescaled according to [41], as studies have shown that track reconstruction in the COT underestimates

[4]Which allows particle identification based on the specific ionization dE/dx, as seen in section 2.2.3.

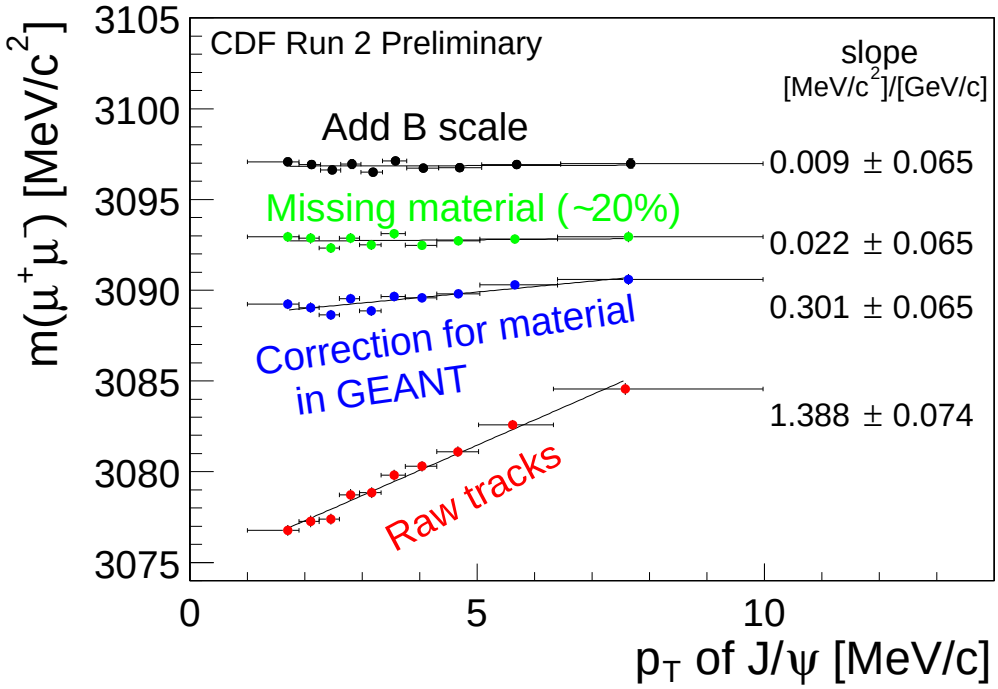


Figure 4.2: The removal of the p_t dependence of the reconstructed invariant J/ψ mass observed when using raw track is done by tuning the material model used for track fitting to compensate the incomplete detector description present in that model. The J/ψ mass is then moved to the world average value by correcting the value of the magnetic field B used for conversion of track curvature to momentum [39].

the measurement errors. The refit starts from the rescaled COT tracks and tries to add silicon information to the track seed as described in the previous section. Hits from the silicon detector subsystems SVX, ISL and L00 are taken into account and the Kalman package [13] is used for track refitting.

4.2.3 Trigger Confirmation

The optimization procedure described in chapter 6 critically depends on a correct description of the properties of exclusive B decays in simulated events. An important aspect of a good description is a precise simulation of the trigger system used to preselect B candidates in the experimental data. In real events involving the decay of B mesons several effects might lead to an event being accepted by the Two Track Trigger (TTT), see 4.1.1.

1. The decay of the B meson produces two or more displaced tracks fulfilling the

trigger requirement.

2. Two random tracks, one potentially resulting from the B decay, the other or both potentially resulting from a background process or multiple interactions in the event, fulfill the trigger requirement.

3. One track from the B mesons decay and an XFT track, that by mistake picked up several wrong hits and so has a tendency towards high impact parameters, fulfill the trigger requirement.

Simulated events are not capable of reproducing the properties of combinatorial background events as the complex environment of hadron collisions is not yet fully understood and known models are not able to reproduce the data. Therefore items 2 and 3 will not be modeled correctly in a simulation. To compensate for this lack of a full background simulation, for each event in real data it is confirmed that indeed two tracks originating from the reconstructed B meson decay were those that fulfilled the trigger requirement and caused the event to be preselected, so the cases 2 and 3 are eliminated in data.
The trigger confirmation is checked by matching reconstructed offline tracks to tracks reconstructed online by the SVT. The SVT information for each event is stored in the SVTD bank and can be accessed for each event. For each B candidate it is then confirmed that at least two tracks in the final state are actually SVT tracks and that at least one pair of SVT tracks from that candidate has passed the Two Track Trigger requirements outlined in 4.1.1.

4.2.4 B Meson Reconstruction

The reconstruction of exclusive B decays is performed using the `BottomMods`[42][43] package. The reconstruction software is highly modular and starts with the reconstruction from simple objects, like reconstructed tracks, to higher level objects, like reconstructed B mesons. The output of each module serves as input to the next module in the reconstruction chain.
The reconstruction starts with a preselection of tracks requiring the track to have a successful helix fit attached to it and requiring a physical error matrix for the track parameters[5] . For each event all reconstructed tracks are refitted with either pion or kaon mass hypothesis and stored internally as collections of stable particle candidates serving as input to the modules reconstructing the decays of unstable particles. In subsequent chapters, two exclusive B decays are of particular interest: the decay of the B^0 meson into the final state $D^-\pi^+$ with the D meson decaying into $K^-\pi^-\pi^+$,

[5]Due to measurement errors and numerical inaccuracies sometimes a negative (unphysical) error is calculated for a track parameter. This occurs very rarely but causes the software to crash.

with the decay of the B_s meson into $D_s^-\pi^+$ and the D_s decaying into the $\phi\pi^-$ and the ϕ decaying into a pair of charged kaons[6]. In the case of the B_s, first ϕ candidates are formed from pairs of oppositely charged tracks assumed to be kaons. Most of those pairs won't originate from a real ϕ decay and are random combinations instead. To reject background candidates, the invariant mass of the two tracks is calculated from the sum of their four momenta. However, as the tracks' four momenta are taken at the perigee point[7], which does not necessarily correspond to the position where the ϕ actually decayed, this is only a rough estimate. A more accurate value for the invariant mass can be obtained only after a full vertex fit. As vertex fits are expensive in terms of computation time, a soft preselection of track pairs based on the estimated invariant mass can make the reconstruction faster. The software performing the vertex fit is based on the `CTVMFT` [44] package, which calculates the vertex positions, invariant masses, four momenta of the particles involved and the corresponding covariance matrices. All track pairs with successful vertex fit and in the mass window from 1.005 to 1.035 GeV around the expected invariant ϕ mass are then accepted as ϕ candidates. In the next step, $D_s \to \phi\pi$ candidates are formed by adding a charged track assumed to be a pion to the ϕ candidate. Candidates with duplicate tracks are removed. A vertex fit is performed on all remaining candidates constraining all three tracks to a common vertex. Candidates with a successful vertex fit and an invariant mass in the mass window from 1.87 to 2.07 GeV, around the expected D_s mass, are used in the reconstruction of the decay $B_s \to D_s\pi$. Here the reconstruction is done in a similar way: tracks from the pion collection are added to all D_s candidates. Candidates with duplicate tracks are removed and B_s candidates are preselected using the four momenta sum of the tracks. At this stage the trigger requirement described in the previous section is applied to the list of B_s candidates. Each candidate is required to include at least one pair of tracks matched to SVT tracks and at the same time fulfilling the Two Track Trigger requirements for SVT track pairs. The fitted decay vertex of the D_s is allowed to be displaced from the B_s vertex due to the lifetime of the D_s of $c\tau = 147\,\mu m$ [15], instead the three momentum of the D_s is constrained to point back to the B vertex. Also the invariant mass of the D_s is constrained to its PDG (particle data group) value of 1.968 GeV [15]. These constraints add additional information to the vertex fit and reduce the number of free parameters, leading to an improved resolution of the reconstructed particle's quantities.
The reconstruction of the decay $B^0 \to D\pi, D \to K\pi\pi$ proceeds similarly. Here as first step one kaon and two pion candidates with the right sign charge combination form a D candidate. After preselection and vertex fit this D collection serves as input to the reconstruction of the decay $B^0 \to D\pi$. Again the trigger requirement is imposed

[6]The corresponding decays of the anti–particles $\bar{B}^0$ and $\bar{B}_s$ are not listed separately but always implicitly included.

[7]The perigee denotes the point of closest approach of a track to the origin of the coordinate system.

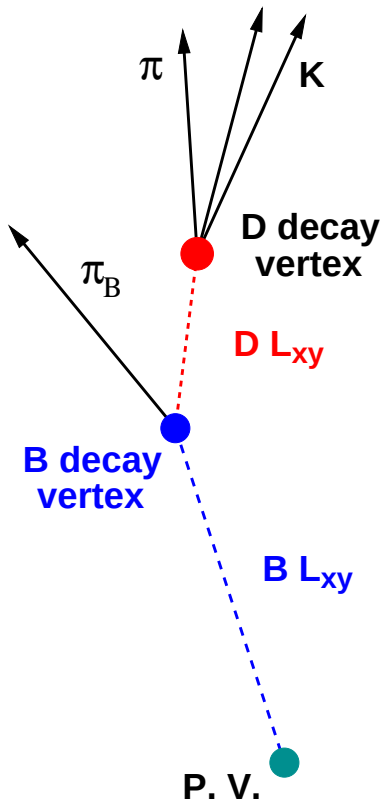


Figure 4.3: Topology of the decay $B^0 \to D\pi, D \to K\pi\pi$ [45].

for each B^0 candidate and the D meson is constrained to its PDG mass and forced to point back to the B vertex. Figure 4.3 shows the typical event topology of a B^0 decay. At each level of the reconstruction, soft preselection cuts are applied to the collections of reconstructed unstable particle candidates to reduce the CPU usage and to reduce the size of the output files. All selection cuts applied to the reconstructed decays described here are listed in table 4.1.

4.2.5 B Stntuples

For storing the information contained in a list of reconstructed B meson candidates, the *B Stntuple* framework was developed [46]. This framework introduces a data format that allows to access all vertex fit informations calculated in the reconstruction chain without having to repeat the fit. When the reconstruction of B meson candidates is completed for a given event, the collections of unstable particles created by successively building up the decay chain as described in the previous section is

Decay	Preselection after vertex fit
$\phi \to KK$	$1.005 < m_\phi[\text{GeV}] < 1.035$
	$\chi^2 < 15$
$D_s \to \phi\pi$	$1.87 < m_{D_s}[\text{GeV}] < 2.07$
	$\chi^2 < 25$
	$L_{xy}/\sigma_{L_{xy}} > 2$
$B_s \to D_s\pi$	$4.4 < m_{B_s}[\text{GeV}] < 6.6$
	$\chi^2 < 25$
	$L_{xy}/\sigma_{L_{xy}} > 2$
$D \to K\pi\pi$	$1.77 < m_D[\text{GeV}] < 1.97$
	$\chi^2 < 25$
	$L_{xy}/\sigma_{L_{xy}} > 4$
$B^0 \to D\pi$	$4.4 < m_{B_s}[\text{GeV}] < 6.6$
	$\chi^2 < 25$
	$L_{xy}/\sigma_{L_{xy}} > 2$

Table 4.1: Preselection cuts applied during the reconstruction of decay candidates. m_p denotes the invariant mass obtained from a vertex fit of particle p, χ^2 the two–dimensional χ^2 obtained from that fit and L_{xy} the two–dimensional decay length (detailed descriptions of all variables can be found in appendix A).

converted into B Stntuple format and stored for later analysis. Figure 4.4 shows the exemplaric data structure of a candidate for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$. For each unstable particle in the decay chain, the results of the vertex fit are stored in a `TDecayPart` object. Stable particles corresponding to track refits with a certain mass hypothesis are stored as `TStablePart` objects, containing all track related informations.

An additional data structure stores general information independent of the actual reconstruction such as run number, event number and online trigger flags for each event. In addition for each B meson candidate the tagging decisions provided by different tagging algorithms (see section 5.4) are stored.

Producing B Stntuples from the Two Track Trigger dataset is a very CPU time consuming task. For each event in the Two Track Trigger dataset, a list of $\simeq 30$ exclusive decay channels of the B^0, $B^\pm$ and B_s mesons is reconstructed, requiring full vertex fits of all mesons and all intermediate resonances involved. To facilitate this process and make efficient use of the resources available at CDF II, the B Stntuple production

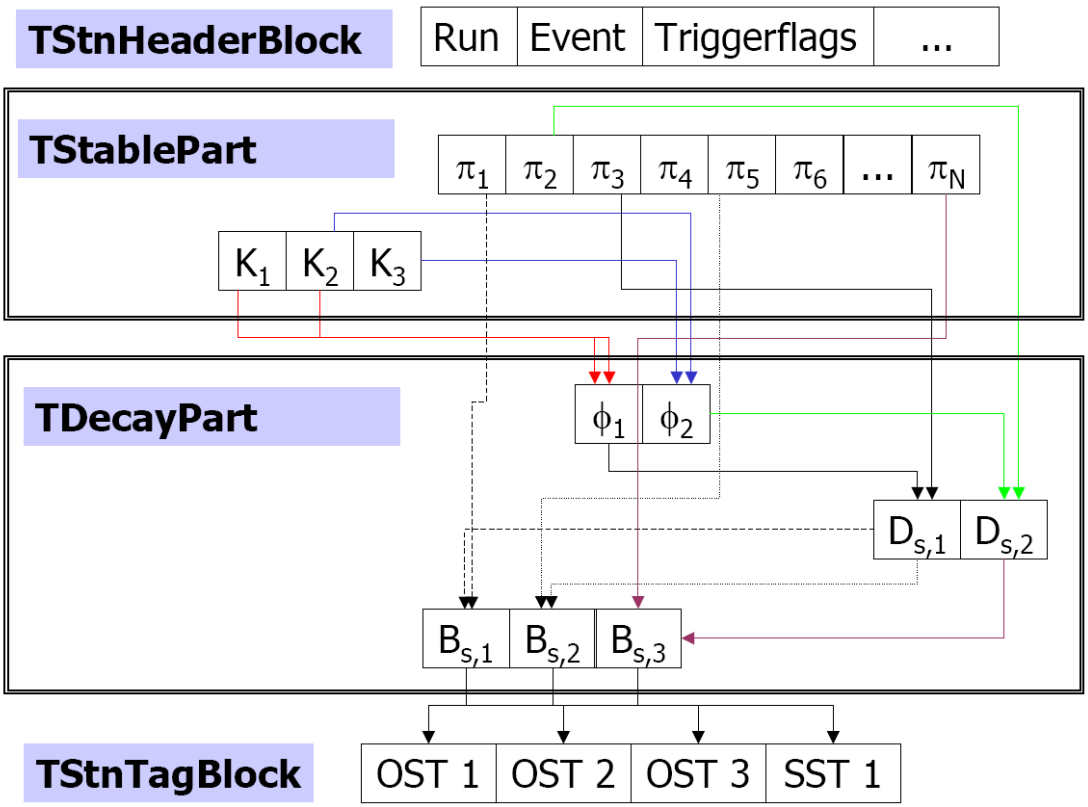


Figure 4.4: Example data structure of an event with three candidates for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ in B Stntuple format.

is centrally organized by the B_s mixing group. The full Two Track Trigger dataset is split into several subsections and each member of the group uses his/her account to submit jobs performing the reconstruction to the CAF (Central Analysis Farm), the main computing resource at CDF II. Using $\simeq$ 30 user accounts, the full B Stntuple production still takes several weeks to process the full 40 TB of Two Track Trigger data. The resulting output B Stntuples are written to mass storage and can be used for later analysis. The B Stntuple dataset resulting from the processing of the Two Track Trigger dataset still amounts to 5 TB, too much to store it on a desktop computer and analyze it conveniently. Further more, in order to read the event structure, the installation of a CDF Software environment is necessary. With the outline to reduce the amount of disk storage space needed and to make iteration over the reconstructed data sample in one particular final state faster and independent of CDF Software, an interface creating small ROOT Ntuples from the B Stntuple datasets was developed as part of this thesis. For each particle in the particular decay chain of

interest only selected quantities are written out to the ROOT Ntuple, still conserving the event structure. This makes the datasets easily available for further analysis using the ROOT framework [47] and reduces the size of the data sample of a particular decay chain reconstructed from the Two Track Trigger dataset to $\mathcal{O}(10\,\mathrm{GB})$.

Chapter 5

Outline of the Mixing Analysis

5.1 Overview

The parameter Δm_s determines the time evolution of the oscillation of a B_s meson into a $\bar{B}_s$ meson and vice versa[1]. Using eq. 1.3 on page 24, one can derive the number of events that were produced as B_s and decayed as B_s (*unmixed* events) or were produced as B_s and decayed as $\bar{B}_s$ (*mixed* events) as

$$N_{unmixed} = \frac{N_0}{2} e^{-\Gamma t}(1 + \cos(\Delta m_s\, t)) \tag{5.1}$$
$$N_{mixed} = \frac{N_0}{2} e^{-\Gamma t}(1 - \cos(\Delta m_s\, t))$$

where Γ denotes the decay width[2] and N_0 the number of B_s mesons produced at $t = 0$. Introducing the time dependent asymmetry A, which expresses the time dependent difference of the number of mixed and unmixed events, one can write

$$A = \frac{N_{unmixed} - N_{mixed}}{N_{unmixed} + N_{mixed}} = \cos(\Delta m_s\, t) \tag{5.2}$$

Several prerequisites are necessary to measure A. First the decay of a B_s meson into a flavor eigenstate has to be reconstructed. As a next step, the flavor of the meson at production time has to be determined by tagging algorithms, or short *taggers*. In general, taggers provide a right decision (right sign tag, or short RS), a wrong decision (wrong sign tag, WS) or no decision at all (no tag, NT). A measure of the quality of a

[1] All formulas apply to the B^0 as well by replacing Δm_s by Δm_d in all expressions.
[2] Here for both B^0 and B_s meson the approximation $\Delta\Gamma_{d,s} = 0$ is used (see section 1.4.4).

tagger is given by the tagging power ϵD^2, which is composed of the tagging efficiency ϵ, defined as

$$\epsilon = \frac{N_{RS} + N_{WS}}{N_{RS} + N_{WS} + N_{NT}} \tag{5.3}$$

and the tagging dilution D, that gives a measure how often the tagger provides a wrong tag:

$$D = \frac{N_{RS} - N_{WS}}{N_{RS} + N_{WS}} \tag{5.4}$$

The dilution D is related to the probability P that a tagging decision correctly identifies the flavor: $D = 2P - 1$. A perfect tag has $D = 1$ and a random tag has $D = 0$. The measured asymmetry is given by $A_{meas} = D\,A_{true}$, hence the "dilution" should be maximized.
For a measurement of time–resolved B_s oscillations, the lifetime of the B_s must be taken into account, as for a given decay time t the number of B_s mesons that have mixed (not mixed) [3] has to be determined. A parametrization of the B_s lifetime distribution has to incorporate the limited experimental resolution of the reconstructed B decay vertex and also model the bias in the lifetime distribution introduced by preselecting events with displaced tracks at trigger level (see section 4.1.1).
In addition, lifetime parametrization and tagging have to take into account that the reconstructed sample of B mesons contains a fraction of background events of several types. The composition of the data sample can be analysed by studying the invariant mass spectrum of the reconstructed B meson. In general, each reconstructed exclusive final state will contain different types of backgrounds and their contributions have to be understood and modeled for each decay channel individually.
In summary, the necessary ingredients for a Δm_s measurement are:

- Understanding the invariant mass spectrum to determine the composition of signal and background events in the sample of B mesons candidates.
- Parametrization of the B meson lifetime distribution for a time–resolved measurement of the oscillations.
- Flavor tagging to determine whether a B meson has mixed or not.

Each individual step is discussed in detail in the following sections. First section 5.2 discusses the composition of the invariant mass spectra of the decays $B^0 \to D\pi, D \to K\pi\pi$ and $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$, which are the main focus of the work presented in this thesis. Section 5.3 then introduces a model describing the time

[3] A B_s meson can have oscillated several times before it decays. Saying it has "mixed" in this context means that the meson has performed an odd number of oscillations, so the flavor eigenstates at production and decay time are different.

evolution of B decays with trigger bias and finite resolution, before in section 5.4 different tagging algorithms used at CDF II are introduced. Section 5.5 then combines mass, lifetime and tagging information to a combined probability density function that can be used to extract Δm_s (or equivalently Δm_d). Finally the sensitivity for a Δm_s measurement is discussed in section 5.6.

5.2 Understanding the Invariant Mass Spectrum

The result of the decay vertex reconstruction procedure described in section 4.2.4 is a set of B meson candidates. Only a fraction of all combinations of tracks that are fitted to a common vertex and pass all preselection cuts will in fact originate from the decay of a B meson into the reconstructed exclusive final state.
Figure 5.1 shows a typical invariant mass spectrum obtained after the vertex reconstruction. Several distinct features can be observed in the spectrum: At $\simeq 5.3$ GeV the signal peak[4], corresponding to the decay of the B meson into the reconstructed exclusive final state is visible. In the lower invariant mass range "satellite peaks" and a significant rise of the spectrum can be observed. These structures can be assigned to partially reconstructed B mesons where one or more, possibly neutral, particle(s) are missed in the reconstruction of the B meson decay. Another feature of the invariant mass spectrum shown in figure 5.1 is a smooth distribution of events spreading over the full invariant mass range. This so–called "combinatorial background" originates from random combinations of tracks that happen to pass the vertex fit requirement and all preselection cuts, but do not result from the decay of a b–hadron. However, in a decay chain reconstructed as $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ for example, these events could still contain the decays of real D_s or ϕ mesons that are combined with one or two random tracks and pass the four track vertex fit requirement. In addition to combinatorial background and partially reconstructed decays, there is also the possibility of reconstructing the exclusive final state of a different species of b–hadrons (which is not the one that one actually tries to reconstruct) where a wrong particle hypothesis is assigned to one or more particles in the decay chain. Examples for this type of background are the decays $\Lambda_b \to \Lambda_c\pi, \Lambda_c \to pK\pi$ and $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ that can be reconstructed as $B^0 \to D\pi, D \to K\pi\pi$, if in the final state the proton originating from the Λ_b decay or one kaon from the B_s decay are by mistake considered as pions. The background contribution from these misassigned particle hypothesis is in general small, and therefore not labeled explicitly in figure 5.1. Nevertheless, these events tend to overlap with the main signal peak, and so have to be taken into account when parametrizing the different background contributions for a mixing measurement.

[4]Note that in order to visualize the signal peak, already strong preselection cuts are applied to the dataset obtained after the vertex reconstruction.

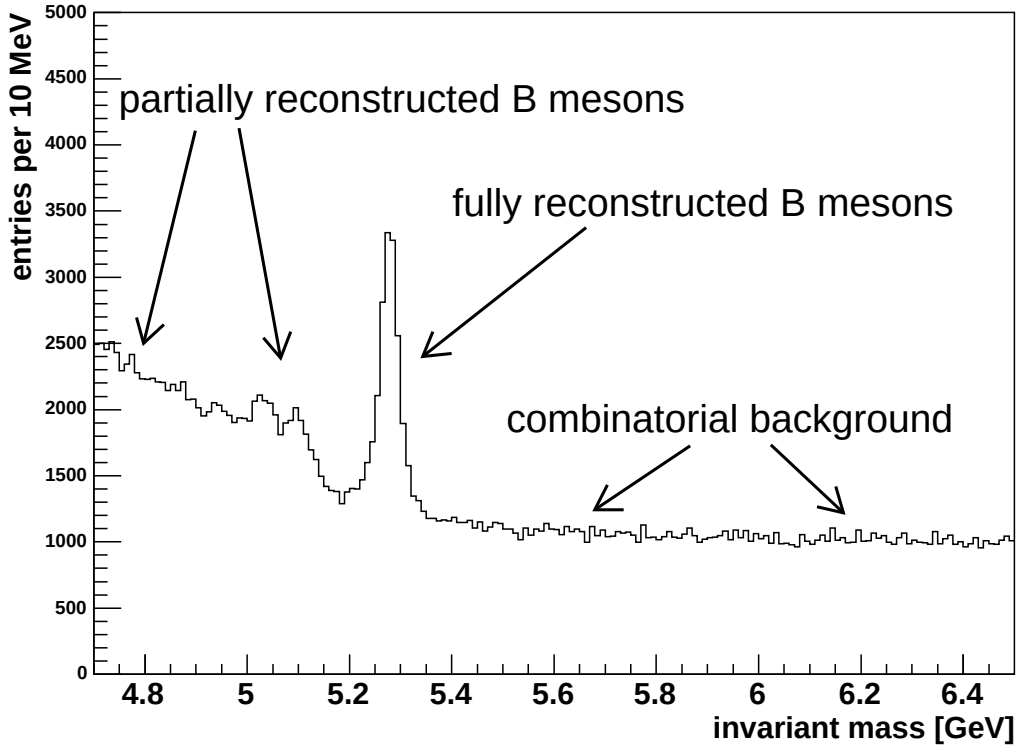


Figure 5.1: Typical shape of the invariant mass spectrum of a reconstructed B meson. In addition to the fully reconstructed final state, there are various "satellite peaks" on the lower invariant mass side from partially reconstructed B mesons. Also the smooth distribution of combinatorial background events can be observed.

In general, the types of background observed when reconstructing the decays of exclusive B mesons can be grouped into three categories:

- Partially reconstructed B meson decays where one or more particle(s) are missed in the final state. These decays are sometimes denoted as satellite peaks, or reflections.

- Fully reconstructed final states of different b–hadrons with wrong particle hypothesis assigned to one particle in the final state.

- Combinatorial background, consisting of track combinations not originating from the decay of a b–hadron.

In the following two sections, the mass spectra of the two decay channels $B^0 \to D\pi, D \to K\pi\pi$ and $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ are studied in more detail and a parametrization of the shape of the invariant mass spectrum is introduced, taking into account all sources of background discussed in this section.

5.2.1 The Mass Spectrum for $B^0 \to D\pi, D \to K\pi\pi$

Partially reconstructed decays

Simulated events of inclusive B decays[5] help to understand the composition of the shape of partially reconstructed decays. In the case of the decay $B^0 \to D\pi, D \to K\pi\pi$ these contributions can be grouped into two classes:

- The decay modes $B^0 \to D\rho, \rho \to \pi\pi^0$ and $B^0 \to D^*\pi, D^* \to D\gamma$ and $D^* \to D\pi^0$ where in either case the neutral particle is missed in the final state. These decays have distinctive shapes and can be separated from other partially reconstructed B decays in the invariant mass spectrum.

- All remaining B^0 decay modes that are not fully reconstructed due to neutral particles in the final state. These decays are denoted $B^0 \to DX$ and result in a smooth shape of the invariant mass spectrum, rising towards lower invariant masses. These decays include semileptonic B meson decays, like $B^0 \to D\mu\nu$ and $B^0 \to De\nu$.

Figure 5.2 shows the invariant mass distributions obtained for partially reconstructed decays after reconstructing events from an inclusive simulation as $B^0 \to D\pi, D \to K\pi\pi$. The two peaks observed for the decays $B^0 \to D^*\pi, D^* \to D\gamma$ and $D^* \to D\pi^0$ (center panel of fig. 5.2) result from the angular distribution of the π^0 originating from the D^* decay. This angular distribution can be understood when considering the spin–parity composition of this particular decay chain:

$$\begin{aligned} B^0 &\to D^* \; \pi \\ J^P = 0^- &\to 1^- \; 0^- \end{aligned}$$

$$\begin{aligned} D^* &\to D \; \pi^0 \\ J^P = 1^- &\to 0^- \; 0^- \end{aligned}$$

[5] *inclusive* here indicates that all known decay channels of the B meson, weighted according to their branching ratios, are included in the simulation.

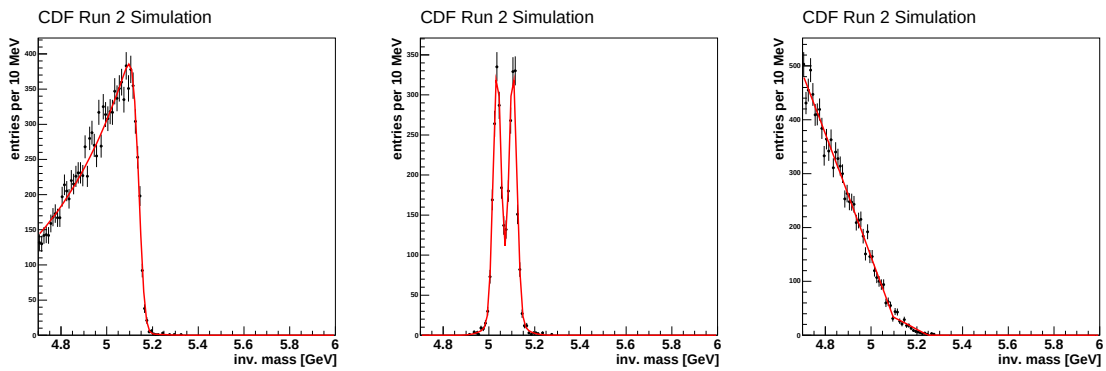


Figure 5.2: Invariant mass distributions resulting from reconstructing the decays $B^0 \to D\rho, \rho \to \pi\pi^0$ (left panel), $B^0 \to D^*\pi, D^* \to D\gamma$ and $D^* \to D\pi^0$ (center panel) and $B^0 \to DX$ (right panel) as the decay $B^0 \to D\pi, D \to K\pi\pi$.

The D^* originating from the B^0 decay is polarized, as, out of three possible spin projections $J_z = -1, 0, 1$, only $J_z = 0$ is allowed due to angular momentum conservation. The angular distribution of the subsequent decay $D^* \to D\pi^0$ then follows a $|d^1_{0,0}|^2 \sim \cos^2\theta$ distribution, where $d^l_{m,m'}$ denotes the corresponding Wigner function [15][48] and θ is the angle between the flight direction of the D^* and the flight direction of the π^0 in the D^*'s rest frame. Therefore the π^0, which is missed in the reconstruction, is emitted mainly (anti–)parallel to the D^*, resulting in the two prominent peaks in the invariant mass distribution of the reconstructed B meson that can be seen in the center panel of figure 5.2.
Figure 5.2 also shows fits of the distributions with parametrized functions that are used as templates describing the shapes of these background contributions when fitting the complete invariant mass spectrum. The function $P_\rho(m)$ chosen to model the shape of the invariant mass distribution of the decay $B^0 \to D\rho, \rho \to \pi\pi^0$ is an exponential function convolved with a Gaussian function plus an additional Gaussian function

$$P_\rho(m) = N_\rho\, e^{-\frac{m}{\tau_1}} \otimes \mathcal{G}(m - m', \mu_1, \sigma_1) + \mathcal{G}(m, \mu_2, \sigma_2) \tag{5.5}$$

Here N_ρ denotes the normalization of the function, μ the mean, σ the width of the Gaussian functions and τ_1 a decay constant. The decays $B^0 \to D^*\pi, D^* \to D\gamma$ and $D^* \to D\pi^0$ are modeled by the sum of three Gaussian functions, where two narrow Gaussian functions are fixed to the same width

$$P_{D^*}(m) = N_{D^*}\left(f_W\,\mathcal{G}(m, \mu_3, \sigma_3) + (1 - f_W)\Big(\mathcal{G}(m, \mu_3 - \xi, \sigma_4) + \mathcal{G}(m, \mu_3 + \xi, \sigma_4)\Big)\right) \tag{5.6}$$

where N_{D^*} is the normalization, f_W the relative contribution of the wide Gaussian function with width σ_3 and mean μ_3 and ξ the relative shift with respect to μ_3 of the

two narrow Gaussian functions with width σ_4. The shape of the remaining partially reconstructed B decays of the type $B^0 \to DX$ is modeled by a combination of an exponential with a linear function

$$P_X(m) = N_X \left(f_L \Big[\frac{2a_1}{(b_1 - m_L)^2} (b_1 - m) \Theta(b_1 - m) \Big] + (1 - f_L) \Big[s_1 \big(1 - (m - m_c)\big) \big(1 - e^{(m-w)\tau}\big) \Big] \right) \tag{5.7}$$

where N_X, a_1 and s_1 are normalization constants, $m_c = m_H - m_L$ denotes the center of the mass fit range, f_L, b_1, w and τ are the fit parameters and $\Theta(b_1 - m)$ is the Heavyside step function.

Misassigned final state particles

Particle identification at CDF II is not powerful enough to decide on a track by track basis whether a particle was a pion, kaon or proton (see section 2.2.3). Hence contributions from other decays with four tracks in the final state, where a kaon or a proton was misidentified as a pion will contribute to the invariant mass spectrum. For the decay $B^0 \to D\pi, D \to K\pi\pi$ such contributions arise from the decays $\Lambda_b \to \Lambda_c\pi, \Lambda_c \to pK\pi$ and $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$. The relative contribution of these decays can be obtained from simulated events. For example, the relative contribution $f_{\Lambda_b \text{ as } B^0}$ of the decay $\Lambda_b \to \Lambda_c\pi, \Lambda_c \to pK\pi$ can be estimated as

$$f_{\Lambda_b \text{ as } B^0} = \frac{\sigma(\Lambda_b)}{\sigma(B^0)} \cdot \frac{BR(\Lambda_b \to \Lambda_c\pi) \cdot BR(\Lambda_c \to pK\pi)}{BR(B^0 \to D\pi) \cdot BR(D \to K\pi\pi)} \cdot \frac{\epsilon(\Lambda_b \text{ as } B^0)}{\epsilon(B^0 \text{ as } B^0)} \tag{5.8}$$

where σ denotes the production cross section, BR the branching ratio of the corresponding decay and ϵ the probability that the corresponding decay will be reconstructed as B^0 by the reconstruction software and pass all preselection cuts. The contribution $f_{B_s \text{ as } B^0}$ of the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$, reconstructed as $B^0 \to D\pi, D \to K\pi\pi$, can be calculated similarly. For this work, the results $f_{\Lambda_b \text{ as } B^0} = 4.1\% \pm 0.8\%$ and $f_{B_s \text{ as } B^0} = 2.0\% \pm 0.4\%$ derived in [49], based on PDG[15] information and recent CDF II publications [50][51] are used.
Another contribution to the invariant mass spectrum results from the Cabibbo suppressed decay $B^0 \to DK$, where the kaon from the B meson decay is reconstructed as a pion. The contribution of this decay relative to the main signal peak $B^0 \to D\pi$ is derived from a simulation and its shape is fixed when fitting the full invariant mass spectrum. The decay $B^0 \to DK$ is considered as an additional signal contribution.
Figure 5.3 displays the shapes and parametrizations of the three decays introduced in this paragraph. The functions $P_{\Lambda_b}(m)$ and $P_{Cab}(m)$ describing the decays $\Lambda_b \to \Lambda_c\pi$

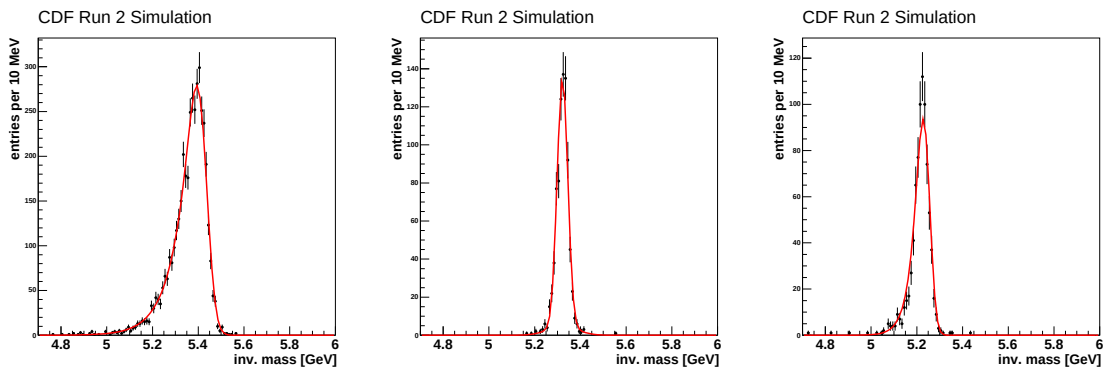


Figure 5.3: Invariant mass distributions resulting from reconstructing the decays $\Lambda_b \to \Lambda_c\pi, \Lambda_c \to pK\pi$ (left panel), $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ (center panel) and $B^0 \to DK$ (right panel) as the decay $B^0 \to D\pi, D \to K\pi\pi$.

and $B^0 \to DK$ are modeled by exponential functions convolved with Gaussian functions, the decay $B_s \to D_s\pi$ is described by $P_{B_s}(m)$, a double Gaussian function with common mean and variable widths.

Fitting the full mass spectrum

The analytic description of the full invariant mass spectrum includes all fit templates derived in the previous two paragraphs. In addition there are functions parametrizing the shape of the invariant mass spectrum of combinatorial background events $P_{comb}(m)$ and a parametrization of the main signal peak $P_{Sig}(m)$. The shape of $P_{comb}(m)$ can't be derived from a simulation, instead an analytical description that parametrizes the shape observed in data is chosen

$$P_{comb}(m) = N\Big((1 - f_o)e^{-\lambda(m-m_c)} + f_o\Big) \tag{5.9}$$

The full expression describing the main signal peak $P_{Sig}(m)$ is chosen as a double Gaussian function with common mean and variable widths. The function $P_S(m)$ includes the main signal peak and all exclusive final states with wrong particle hypothesis in the final state. The relative contributions of these decays are fixed relative to the main signal peak

$$\begin{aligned} P_S(m) \;=\; & (1 - f_{B_s} - f_{\Lambda_b} - f_{Cab})\, P_{Sig}(m) \;+ \\ & f_{B_s}\, P_{B_s}(m) \;+\; f_{\Lambda_b}\, P_{\Lambda_b}(m) \;+\; f_{Cab}\, P_{Cab}(m) \end{aligned} \tag{5.10}$$

The contributions of all partially reconstructed decays and the combinatorial background are merged in $P_B(m)$

$$P_B(m) = (1-f_{part})\,P_{comb}(m) + f_{part}\Big[(1-f_X)\Big((1-f_{D^*})\,P_\rho(m) + f_{D^*}\,P_{D^*}(m)\Big) + f_X\,P_X(m)\Big] \tag{5.11}$$

$P_S(m)$ and $P_B(m)$ are both normalized to one. The fit function $P(m)$ describing the full invariant mass spectrum is given by

$$P(m) = N\Big((1-f_b)\,P_S(m) + f_b\,P_B(m)\Big) \tag{5.12}$$

and combines all individual contributions.
The procedure applied to fit the full invariant mass spectrum starts with binned likelihood fits of the individual templates describing partially reconstructed B^0 decays and exclusive final states with wrong particle hypotheses assigned to one particle on the invariant mass distributions obtained from the inclusive simulation. The corresponding distributions and fits of the templates P_{B_s}, P_{Λ_b}, P_{Cab}, P_ρ, P_{D^*} and P_X are displayed in figs. 5.2 and 5.3. The ratio of the contribution f_{Cab} of the decay $B^0 \to DK$ relative to $B^0 \to D\pi$, the ratio f_X of the decays $B^0 \to DX$ relative to the sum of the decays $B^0 \to D\rho, \rho \to \pi\pi^0$ and $B^0 \to D^*\pi, D^* \to D\gamma$ and $D^* \to D\pi^0$ and the ratio f_{D^*} of the decays $B^0 \to D^*\pi, D^* \to D\gamma$ and $D^* \to D\pi^0$ relative to $B^0 \to D\rho, \rho \to \pi\pi^0$ are fixed to the values derived from the inclusive simulation. f_{B_s} and f_{Λ_b} are fixed to the values derived in the previous paragraph.
The free parameters used in the binned likelihood fit of $P(m)$, the function parametrizing the full invariant mass spectrum observed in real data, are:

- f_{part}, the fraction of partially reconstructed B decays relative to combinatorial background events,
- the background fraction f_b, denoting the ratio of background events (described by P_B) relative to signal events (modeled by P_S),
- all parameters of P_{Sig} and P_{comb}, the functions describing the signal peak and the combinatorial background shape respectively,
- the overall normalization N.

Figure 5.4 shows $P(m)$, fitted in the invariant mass range from 4.7 to 6.0 GeV on a data sample of the reconstructed decay $B^0 \to D\pi, D \to K\pi\pi$. The fit function describes the shape of the invariant mass spectrum well. All individual components adding up to $P(m)$ are displayed. These fits are very important for understanding the

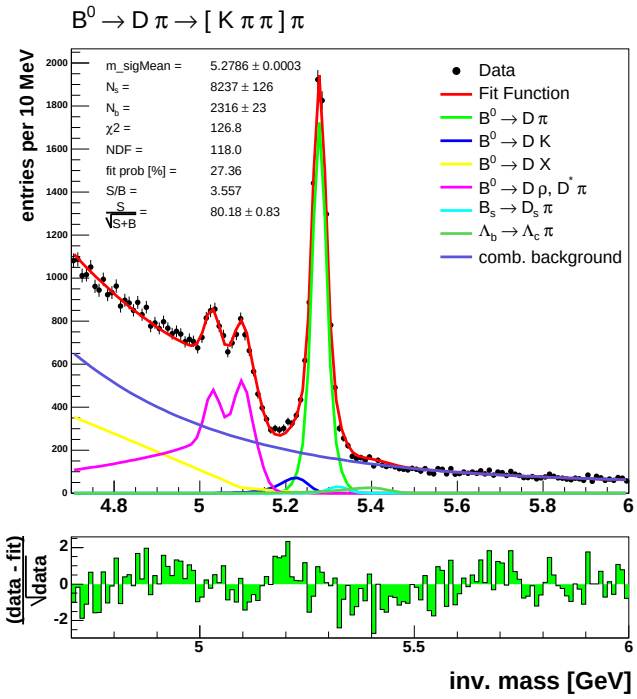


Figure 5.4: Fit of $P(m)$ on the invariant mass distribution obtained for the decay $B^0 \to D\pi, D \to K\pi\pi$ in data.

composition of the data sample used for lifetime and mixing fits, as well as to extracting the number of signal and background events to calculate the signal significance $\frac{S}{\sqrt{S+B}}$. The analytic description of the full invariant mass spectrum is an important ingredient for a fit of Δm_s. As the fit functions for signal and background show a distinct shape and can be easily separated in the mass fit, the fit in invariant mass space provides a handle to separate the relative signal and background contributions in the data sample. When fitting for Δm_s this is taken into account by performing the unbinned maximum likelihood fit simultaneously in mass and lifetime space and using the background fraction f_B as a free parameter, shared by both spaces.

5.2.2 The Mass Spectrum for $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$

Basically everything laid out in the previous section for the decay $B^0 \to D\pi, D \to K\pi\pi$ is also valid for the invariant mass spectrum of the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$. However, the number of reconstructed B_s decays at CDF II is significantly smaller compared to the number of B^0 events. Therefore the contributions from

partially reconstructed decays $B_s \to D_s X$, $B_s \to D_s \rho, \rho \to \pi\pi^0$ and $B_s \to D_s^* \pi, D_s^* \to D_s \gamma$ and $D_s^* \to D_s \pi^0$ can be modeled by a single template $P_{refl}(m)$, consisting of a linear function plus a Gaussian function

$$P_{refl}(m) = N\left((1-f_G)\Big[\frac{2a_2}{(b_2-m_L)^2}(b_2-m)\Theta(b_2-m)\Big] + f_G\,\mathcal{G}(m,\mu_5,\sigma_5)\right) \quad (5.13)$$

The template $P_{Cab}(m)$ describing the decay $B_s \to D_s K$ is modeled by a single Gaussian function. Contributions from the decays $\Lambda_b \to \Lambda_c \pi, \Lambda_c \to pK\pi$ and $B^0 \to D\pi, D \to K\pi\pi$ reconstructed as $B_s \to D_s \pi, D_s \to \phi\pi, \phi \to KK$ with wrong particle hypothesis in the final state can be almost completely removed by requiring a cut on the invariant mass of the decay $\phi \to KK$ and are therefore neglected in the fit. The function $P_S(m)$ describing the main signal peak is chosen to be a single Gaussian function and the combinatorial background is modeled in analogy to equation 5.9 as an exponential function plus offset. The fit function $P(m)$ used for the fit of the full invariant mass spectrum can be written as

$$P(m) = N\left((1-f_b)\,P_S(m) + f_b\left[(1-f_{refl})\,P_{comb}(m) + f_{refl}\,P_{refl}(m)\right]\right) \quad (5.14)$$

where

$$P_S(m) = (1-f_{Cab})\,P_{Sig}(m) + f_{Cab}\,P_{Cab}(m) \quad (5.15)$$

Figure 5.5 shows the invariant mass spectrum of the decay $B_s \to D_s \pi, D_s \to \phi\pi, \phi \to KK$ (left) and the corresponding contributions resulting from partially reconstructed decays modeled by P_{refl} (right).

5.3 Measuring the B Meson Lifetime

5.3.1 Introduction

A situation frequently occurring in quantitative science is the measurement of a distribution of the physical quantity x. A set of measurements x can be regarded as a sample from a probability density function $f(x)$. The true distribution $f(x)$ differs from the measured distribution because of statistical fluctuation and the imperfect measurement process that prevents a measurement of the true value of x. Examples for such experimental limitations are a limited acceptance and the finite precision of the measurement.

One can express this interrelationship formally [33]: Instead of x, in reality the quantity y is measured and as the result of a set of measurements a probability density

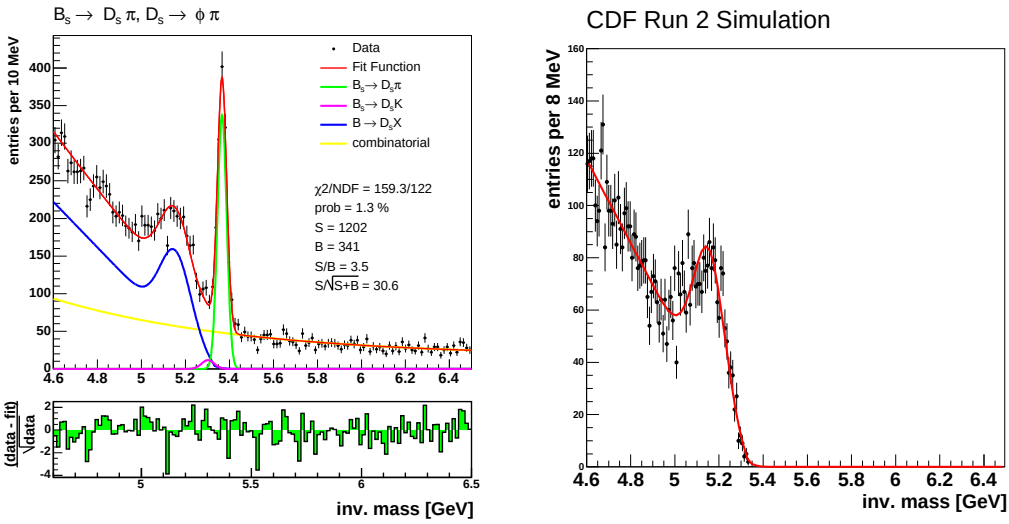


Figure 5.5: Fit of $P(m)$ on the invariant mass distribution obtained for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ in data (left). The right panel shows the corresponding template describing the contributions from the decays $B_s \to D_s X$, $B_s \to D_s\rho, \rho \to \pi\pi^0$ and $B_s \to D_s^*\pi, D_s^* \to D_s\gamma$ and $D_s^* \to D_s\pi^0$ to the invariant mass spectrum derived from simulated events.

function $g(y)$ in this quantity is obtained. The distribution $f(x)$ of the physical variable x defined in $a \geq x \geq b$ and the measured distribution $g(y)$ of the quantity y are connected via the convolution integral

$$g(y) = \int_a^b A(y,x)\, f(x)\, dx \tag{5.16}$$

Here the kernel $A(y,x)$ of the integral equation determines the probability of measuring a value y when the true value of the physical quantity is x. This corresponds to the so–called *Fredholm Integral Equation of the First Kind.*

The determination of the lifetime of a fully reconstructed B meson with data collected from the two track trigger at CDF II is an example for such a task. The observable t, the proper decay time of the B meson, is composed of measured quantities of the reconstructed B meson decay vertex and is given by

$$t = \frac{L_{xy}\, m_B^{\mathrm{PDG}}}{c\, p_t} \tag{5.17}$$

where L_{xy} denotes the decay length, p_t the transverse momentum of the reconstructed B meson candidate and c the velocity of light. m_B^{PDG} corresponds to the current

world average value of the invariant mass of the corresponding B meson from [15]. The precision of the measurement of t is limited by the experimental decay length resolution $\sigma_{L_{xy}}$ of the B vertex reconstruction and the resolution of the measurement of the transverse momentum p_t.

$$\sigma_t = \sqrt{\left(\frac{m_B^{\mathrm{PDG}}}{c\,p_t}\sigma_{L_{xy}}\right)^2 + \left(\frac{m_B^{\mathrm{PDG}} L_{xy}}{c\,p_t^2}\sigma_{p_t}\right)^2} = \sqrt{\left(\frac{m_B^{\mathrm{PDG}}}{c\,p_t}\sigma_{L_{xy}}\right)^2 + \left(\frac{t}{p_t}\sigma_{p_t}\right)^2} \tag{5.18}$$

The momentum resolution σ_{p_t} for individual tracks is extremely good (see section 4.2.1). For semileptonic decays the missing neutrino momentum, which is corrected for with an average correction factor taken from simulations, introduces an additional uncertainty in the p_t of the reconstructed B meson. For exclusive hadronic final states the contribution of σ_{p_t} to σ_t is negligible compared to $\sigma_{L_{xy}}$. This simplifies eq. 5.18 to

$$\sigma_t = \frac{\sigma_{L_{xy}}\; m_B^{\mathrm{PDG}}}{c\,p_t} \tag{5.19}$$

The acceptance for measuring different values of t is sculpted by two effects: Already at trigger level the shape of the measured proper decay time distribution is modified by preselecting events that contain at least two SVT tracks displaced from the primary vertex. A cut on SVT tracks requiring $100\,\mu m < d_0 < 1\,mm$ for two tracks in the event is applied, resulting in a reduced acceptance for both short– and long–lived B mesons. In addition, selection cuts to improve the signal to noise ratio of the observed B meson resonance using cuts on the decay length, impact parameter or also cuts on a neural network output can additionally sculpt the shape of the observed proper decay time distribution.
In general the kernel function $A(t, t_{\mathrm{true}})$ is a function of both the unknown true lifetime t_{true} and the measured lifetime t, incorporating at the same time resolution and acceptance effects. The determination of A is a difficult task and would require an unfolding procedure. Analyses at CDF II requiring the reconstruction of the B meson lifetime distribution taking into account a limited acceptance introduced by trigger and analysis cuts make use of a simplified model (see for example [49][52][53][54]) laid out in the following section.

5.3.2 Probability Density Function in Lifetime Space

Ideally one would expect the probability $P_S(t)$ to observe the decay of a particle with lifetime τ at a proper decay time t to follows an exponential decay law

$$P_S^{\mathrm{unbiased}}(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \tag{5.20}$$

As discussed in the previous section, this distribution is affected by the limited experimental resolution σ_t and the limited acceptance that is sculpted by trigger preselection and analysis cuts. As an approximation of equation 5.16, these effects can be modeled as a function of the measured lifetime t by the product ansatz [53]:

$$P_S^{\text{biased}}(t, \sigma_t) \equiv P_S(t, \sigma_t) = \widetilde{P}_S(t, \sigma_t) \cdot \epsilon(t) \tag{5.21}$$

where

$$\widetilde{P}_S(t, \sigma_t) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes \mathcal{G}(t - t', \sigma_t) \tag{5.22}$$

and $\otimes$ denotes the convolution integral

$$\frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes \mathcal{G}(t - t', \sigma_t) = \int_0^\infty \frac{1}{\tau} e^{-\frac{t}{\tau}} \mathcal{G}(t - t', \sigma_t)\, dt' \tag{5.23}$$

of the unbiased lifetime distribution with a Gaussian function $\mathcal{G}$ centered at zero and with width σ_t, modeling the finite resolution of the proper decay time measurement. In general σ_t has a different value $\sigma_{t,i}$ for each event i in the data sample.
The efficiency function $\epsilon(t)$ used in eq. 5.21 is used to take into account the limited acceptance introduced by trigger and cut selections. $\epsilon(t)$ can be determined by comparing the signal proper decay time distribution $g(t)$ *after* all trigger and analysis cuts are applied to an unbiased lifetime distribution (eq. 5.20), convolved with a resolution function taking into account the event–by–event uncertainty of the proper decay time measurement $\sigma_{t,i}$ of the vertex position [54]:

$$\epsilon(t) = \frac{g(t) \text{ after trigger and cuts}}{\sum_i \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes \mathcal{G}(t - t', \sigma_{t,i})} \tag{5.24}$$

Here i sums over all events in the sample contained in $g(t)$. The biased signal distribution $g(t)$ can be obtained from a simulation of B meson decays, properly taking into account the effects of trigger preselection.

5.4 Flavor Tagging

An important ingredient to a measurement of time–resolved B_s oscillations is the information whether a B mesons has mixed or not. Therefore the flavor, or quark content, of the B meson at production and at decay time has to be determined. For a fully reconstructed decay into a flavor eigenstate, the flavor of a B meson at decay time is determined by its decay products. To decide whether this particle has mixed or not, still the flavor of the B meson at production time has to be determined using flavor tagging algorithms.

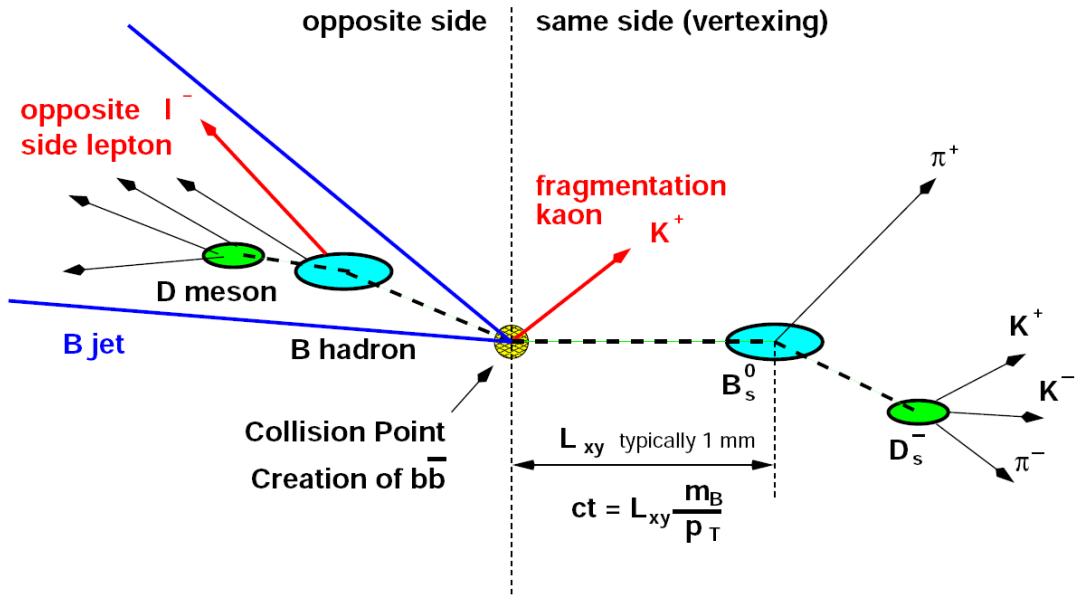


Figure 5.6: Visualization of different flavor tagging algorithms.

Flavor tagging algorithms can generally be classified into *opposite side* (OST) and *same side* (SST) tagging algorithms. Opposite side tagging exploits the fact that b quarks are produced in $p\bar{p}$ collisions as pairs of a b and $\bar{b}$ quark. This can be used to determine the flavor of the b quark that lead to the production of the reconstructed B meson by determining the flavor of the other b quark in the event. Same side flavor tagging relies on the charge of particles produced in association with the production of the reconstructed B meson.

Currently three opposite side tagging algorithms are used at CDF II: soft muon, soft electron and jet charge tagging. The soft lepton tags [55] are based on semileptonic B meson decays. The charge of the lepton indicates the flavor of the b quark: negatively (positively) charged leptons come from direct b ($\bar{b}$) decays. Misidentification of the b flavor can arise from leptons produced in subsequent D meson decays or from fake leptons, where a pion or kaon is misidentified as a lepton. The efficiency of lepton flavor tagging is limited by the small branching fraction of semileptonic B mesons decays and the efficiency of lepton identification in the detector.

The jet charge tag [56] is based on the momentum weighted sum of charge of tracks associated with a jet produced by the decay of a b quark. Wrong jet charge tags can arise from jets (or single particles in the jet) that don't actually origin from the b quark and from particles in the jet that are not detected.

The SST algorithm used at CDF II is based on the associated particle production during the fragmentation of a b quark to B (or B^{**}) mesons [57]. In the case of B_s mesons this means that the fragmentation of a $\bar{b}$ (b) quark to a B_s ($\bar{B}_s$) meson, containing an s ($\bar{s}$) quark, enforces the production of a $\bar{s}$ (s) quark from the vacuum, preferably leading to the production of a K^+ (K^-) meson. If a charged kaon can be identified nearby a B_s meson, this gives an indication of the flavor of the B_s at production time. The SST Kaon tagging algorithm implemented at CDF II uses the particle identification systems introduced in section 2.2.3 and tries to identify tracks with a high probability to be a kaon in the vicinity of a B_s meson.
Figure 5.6 summarizes the functional principle of the different tagging algorithms visually.

5.5 Extracting Δm_d and Δm_s

A probability density function describing the time dependent evolution of a $B - \bar{B}$ oscillation signal is composed of several parts:

- The asymmetry term including the $\cos(\Delta m\, t)$–modulation of the mixing signal (cf. eq. 5.2). This term includes the tagger decision and has to take into account that the tagger might sometimes give a wrong decision by incorporating the dilution D.
- The lifetime term, describing the decay of the B mesons while they oscillate. This has to take into account the bias introduced by trigger and preselection as laid out in section 5.3.
- The description of signal and background shapes in mass space, as this provides a handle to determine the relative fraction of signal and background contained in the data sample (cf. section 5.2).

Denoting the tagger decision as $\xi = \{-1, 0, 1\}$ for { *mixed, no decision, not mixed* }, the time evolution of the mixing signal can be written as

$$P_S(t, \xi, \sigma_t) = \frac{1}{N_S(t, \xi, \sigma_t)} \cdot \left(\frac{1 + \xi D \cos(\Delta m\, t)}{1 + |\xi|} \frac{1}{\tau} e^{-\frac{t}{\tau}} \right) \otimes \mathcal{G}(t - t', \sigma_t) \cdot \epsilon(t) \qquad (5.25)$$

where D denotes the dilution of the tag decision and $\epsilon(t)$ the lifetime efficiency function introduced in the previous section. $N_S(t, \xi, \sigma_t)$ keeps the function normalized to one. The full analytical expression for $P_S(t, \xi, \sigma_t)$ and the computation of $N_S(t, \xi, \sigma_t)$ can be found in appendix C. It is particularly useful to derive an analytical solution for the normalization of the probability density function, as this has to be calculated for each event individually in the unbinned maximum likelihood fit, as for each event a

different $\sigma_{\tau,i}$ enters eq. 5.25.
The probability density function, incorporating in addition the description of signal and background events in mass space, is given by

$$P(m,t,\xi,\sigma_t) = (1-f_B)\left[P_S(m)\cdot\frac{1}{N_S(t,\xi,\sigma_t)}\cdot\left(\frac{1+\xi D\cos(\Delta m\, t)}{1+|\xi|}\,\frac{1}{\tau}e^{-\frac{t}{\tau}}\right)\right.$$
$$\left.\otimes\,\mathcal{G}(t-t',\sigma_t)\cdot\epsilon(t)\right] + f_B\,P_B(m)\,P_B(t) \tag{5.26}$$

Here f_B is the relative fraction of background events in the data sample and $P_S(m)$ and $P_B(m)$ denote the parametrizations of the invariant mass distributions of signal and background events encountered in section 5.2. $P_B(t)$ is a parametric description of the background's lifetime distribution derived from simulated events (see section 7.2 and appendix E).
During the unbinned maximum likelihood fit for each event

$$\int_m\int_t P(m,t,\xi,\sigma_t)\,dm\,dt = 1 \qquad \text{for } \xi = -1,0,1 \tag{5.27}$$

has to be assured.

5.5.1 Amplitude Scan

For a measurement of Δm expression 5.26 can be used to perform an unbinned maximum likelihood fit on a data sample containing B^0 (B_s) events and use Δm as a free parameter in the fit. If a small number of events is available only, in general the unbinned maximum likelihood fit will not converge and the derivation of limits for Δm will be difficult. An alternative approach, inspired by the idea of performing a Fourier transformation of the oscillation signal, is the *amplitude scan* method, suggested in [12]. This method can also be used to combine results from different experiments directly (for examples see [10]).
Instead of fitting Δm, a scale factor A, the *amplitude*, is introduced in the asymmetry term of the probability density function

$$\frac{1+\xi D\cos(\Delta m\, t)}{1+|\xi|} \longrightarrow \frac{1+\xi AD\cos(\Delta m\, t)}{1+|\xi|} \tag{5.28}$$

This amplitude serves as a free parameter in unbinned maximum likelihood fits using different fixed values of Δm. Ideally the amplitude would come out to be $A = 1$ for the correct value of Δm and compatible with zero for all other values of Δm. Figure 5.7 shows the amplitude scan from [11] as an example.
In figure 5.7 the fit result for the amplitude A is consistent with zero for oscillation

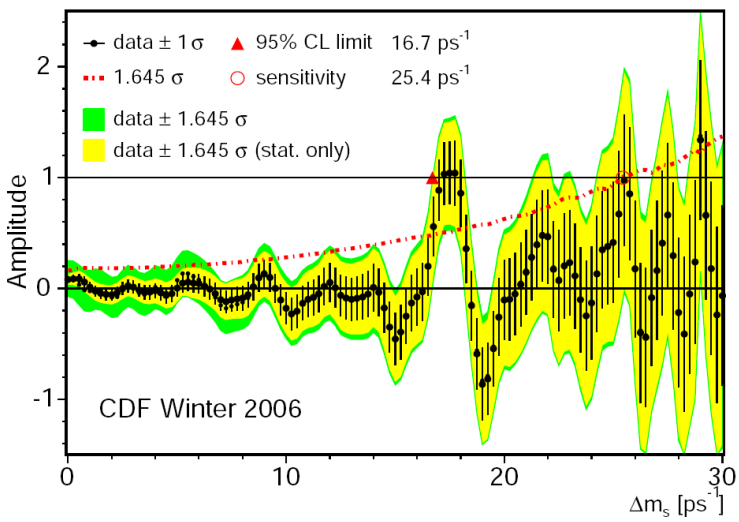


Figure 5.7: Amplitude scan for Δm_s from [11].

frequencies from 0 to 16 ps^{-1}, this means these values of Δm_s can be excluded. The amplitude then rises to one for values of Δm_s from 17 to 17.5 ps^{-1}, which is interpreted as the 3.5 σ signal for B_s oscillations. The amplitude then goes to zero again at $\Delta m_s \simeq 18$ ps^{-1} and is then consistent with zero again for higher frequencies with an increasingly larger uncertainty of A. The sensitivity for different values of Δm_s can be estimated by the *blind amplitude scan* method. In this method the tagging decisions are randomized for the amplitude scan. The frequency at which the 95 % confidence limit of the fitted value of A includes the value $A = 1$ gives the frequency where in average the 95 % exclusion limit on Δm_s can be set in the absence of an oscillation signal if the real tagging decisions are included again in the amplitude scan. In figure 5.7 the result of this blind amplitude scan is indicated by the red dotted line that intersects the line $A = 1$ at a value of $\Delta m_s = 25.4$ ps^{-1}, marked by the red circle. When including the tagging decisions, the 95 % confidence limit of A hits the line $A = 1$ at $\Delta m_s = 16.7$ ps^{-1}, which means that values of $\Delta m_s < 16.7$ ps^{-1} can be excluded with 95 % confidence by the current CDF II measurement [11]. The value of $\Delta m_s = 17.33^{+0.42}_{-0.21} \pm 0.07$ ps^{-1} is not derived from the amplitude scan but obtained after a separate unbinned maximum likelihood fit.

5.6 Sensitivity for Δm_s

The significance of the Δm_s signal observed at $\Delta m_s = 17.33^{+0.42}_{-0.21} \pm 0.07$ ps^{-1} [11] by CDF II is 3.5σ and the probability of such a signal originating from a background

fluctuation is still 0.5 %. To establish a more significant signal and to finally exclude the possibility of a background fluctuation in an updated measurement of Δm_s, it is worthwhile to review the experimental parameters that determine the significance of the Δm_s measurement.
The amplitude of the mixing signal is attenuated by two contributions: the tagging performance and the resolution of the proper decay time distribution. In addition the sensitivity for measuring Δm_s depends on the absolute number of reconstructed signal events and on the purity of the signal, i.e. how effective signal events are selected while background events are suppressed.

5.6.1 Tagging Performance

There are several flavor tagging algorithms used at CDF II to determine the quark content of the reconstructed B meson at production time (cf. section 5.4). Two factors determine the performance of each tagging algorithm – its tagging efficiency ϵ, i.e. how often the tagger gives a decision, and its dilution D, quantifying how often this is a correct decision. Both quantities have already been introduced in section 5.1. The following table summarizes the values measured of the tagging power ϵD^2 for the four different tagging algorithms currently used at CDF II [58].

Algorithm	ϵD^2 [%]
OST Soft Electron	0.3
OST Soft Muon	0.5
OST Jet Charge	0.7
SST Kaon	4.0

5.6.2 Proper Decay Time Resolution

The Lorentz–invariant proper decay time t of particles can be reconstructed from the transverse decay length L_{xy} and the Lorentz boost p_t/m, where p_t denotes the transverse momentum and m the invariant mass of the particle. The uncertainty σ_t on this measurement (see section 5.3) is given by

$$\sigma_t = \frac{\sigma_{L_{xy}}\, m_B^{\mathrm{PDG}}}{c\, p_t}$$

The value of σ_t derived from the vertex fit algorithm (see section 4.2.4, page 53) underestimates the proper decay time resolution by a factor 1.2–1.4 [59]. A rescaling procedure based on reconstructed prompt events is applied to obtain a better estimate of the true resolution of t for exclusively reconstructed B mesons [59]. Figure 5.8

compares the time dependent mixing asymmetries and decay rates introduced in the previous section for $B_s - \bar{B}_s$ and $B^0 - \bar{B}^0$ oscillations. These plots show an idealized situation without taking into account attenuation of the signal by detector effects like resolution and acceptance. It is evident that the proper decay time resolution for the observation of B_s oscillation is much more important than in the case of B^0 oscillations. The rapid oscillations of the B_s meson are smeared by the finite resolution and if the vertex resolution was not sufficiently good, the oscillations would be averaged out and could not be observed any more.

5.6.3 Signal Significance

Expressions for the significance $\mathcal{S}$, i.e. the ratio of the expected amplitude of the oscillation signal over background, can be derived using Fourier and Likelihood methods [12]. Neglecting resolution effects, $\mathcal{S}$ can be expressed as

$$\mathcal{S} = \frac{S}{\sqrt{S+B}} \sqrt{\frac{\epsilon D^2}{2}} \tag{5.29}$$

where S denotes the number of signal events in the reconstructed final state and B the number of background events in the signal region. Taking into account the effect of the limited proper decay resolution on an oscillation signal $P(t)$ by a convolution with a Gaussian function

$$P(t') \otimes \mathcal{G}(t - t', \sigma_t) \tag{5.30}$$

introduces an additional attenuation factor

$$\mathcal{S} = \frac{S}{\sqrt{S+B}} \sqrt{\frac{\epsilon D^2}{2}}\, e^{-\sigma_t^2 \Delta m_s^2/2} \tag{5.31}$$

To improve the measurement of Δm_s, each of these three contributions to the signal significance has to be improved. Particular emphasis will be put on the optimization of the factor $\frac{S}{\sqrt{S+B}}$ in the context of this work.

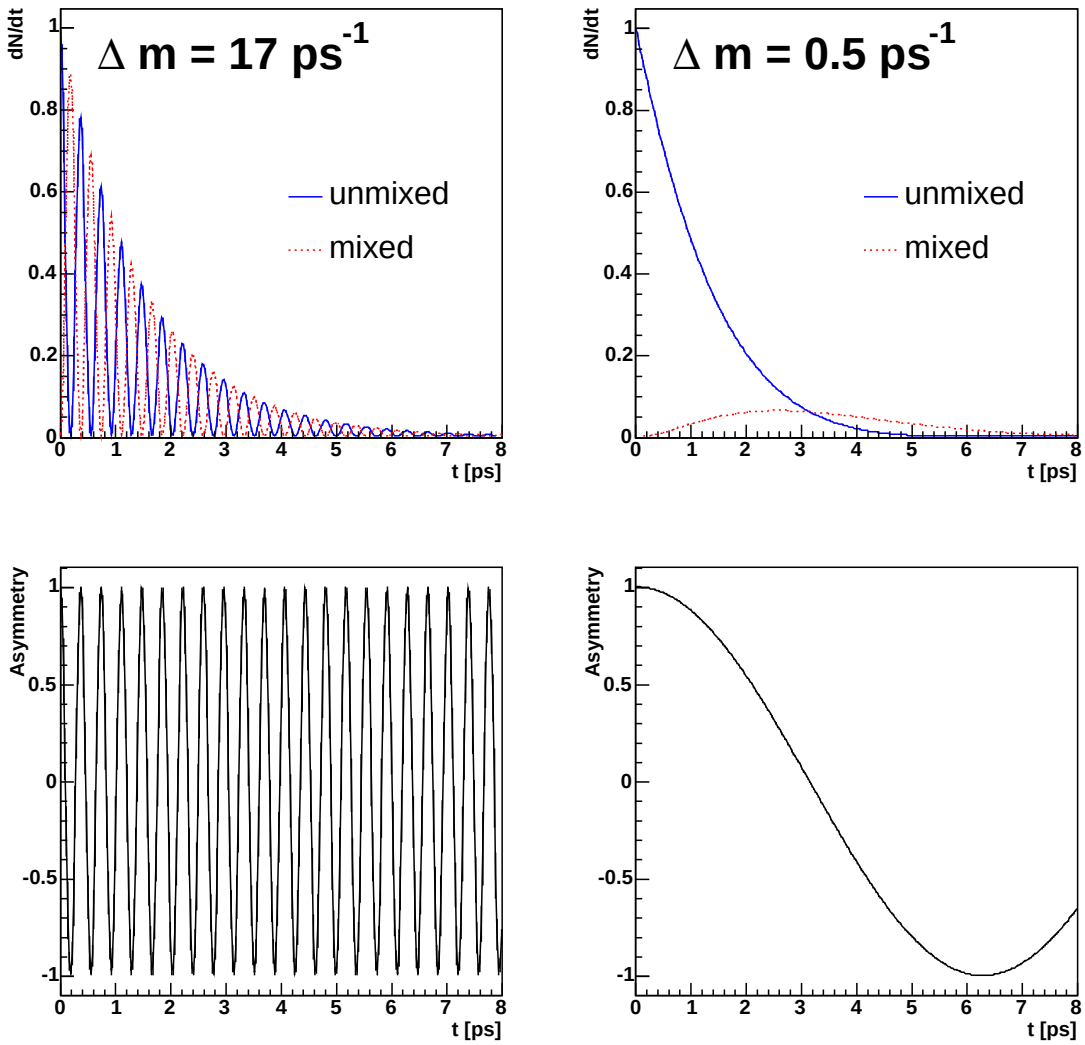


Figure 5.8: Comparison of the time dependent asymmetries and decay rates for $B_s - \bar{B}_s$ oscillations (left plots) and $B^0 - \bar{B}^0$ oscillations (right plots). A mixing frequency of $\Delta m_s = 17$ ps^{-1} was assumed for the B_s and Δm_d was set to 0.5 ps^{-1}. For both states a lifetime $\tau = 1/\Gamma = 1.5$ ps was assumed. No experimental effects like detector resolution or acceptance are taken into account.

Chapter 6

Signal Optimization with Neural Networks

6.1 Motivation

The signal significance $\frac{S}{\sqrt{S+B}}$ is one of three factors contributing to the significance of a B_s mixing signal (cf. eq. 5.31, page 78). S denotes the number of signal events that are obtained at a background level of B events. The method used at CDF II to find the optimum value for $\frac{S}{\sqrt{S+B}}$ up to now employs an iterative cut optimization procedure. This method is based on the combination of cuts on several quantities that separate signal events from background events in the data sample obtained after the vertex reconstruction (section 4.2). The cut applied on each single quantity is varied while the cuts applied to all other quantities are kept fixed. For each cut, the resulting significance $\frac{S}{\sqrt{S+B}}$ is calculated from a fit on the resulting invariant mass spectrum. The cut yielding the maximum value is selected and kept fixed. This procedure is repeated for all selected variables until the set of cuts is obtained that maximizes $\frac{S}{\sqrt{S+B}}$. A more detailed description of the method can be found in [60] for example.
The cut based optimization procedure does not combine the information provided by the selection variables in an optimal way. Typical quantities that are used to distinguish B meson signal from background events are the decay length L_{xy}, the impact parameter d_0 or the transverse momentum p_t of the candidate. Some of these variables are correlated in the sense that a certain value of one variable will already contain information on the value of other variables. If, for example, the χ^2 of the D meson vertex fit is very small, also the χ^2 of the B meson vertex fit, where a pion and the D meson are fitted to one vertex, will have a tendency towards smaller χ^2 values.
The correlation between variables is not taken into account by the cut based opti-

mization procedure. A neural network is able to exploit to additional information contained in the correlation between variables and optimally combines them into a single discriminating variable.

6.2 Strategy

The training of a neural network for signal optimization requires input or training patterns describing the properties of signal and background events. A simulation can be used to provide input patterns for signal events, assuring it properly describes the distributions of the training variables of signal events in real data. As the main benefit from using a neural network, compared to a sequential cut based selection, arises from taking into account correlations between the training variables, also a correct simulation of these correlations is needed.
As can be deferred from figure 5.4 (68), the composition of the background events in the signal region is mainly dominated by combinatorial background, i.e. events which contain either a random combination of four tracks, or events which include the decay of a real D meson that is combined with a random track not originating from the decay of a B meson. To reduce the number of background events in the signal region it is therefore desirable to suppress such combinatorial background events.
To obtain a simulation of signal events, only the resonance decays of B mesons into exclusive final states have to be simulated, which makes the simulation ideally suited to provide training patterns for signal events to the network training. To provide combinatorial background events, the full fragmentation process of quarks to hadrons has to be simulated. The complex environment of hadron collisions is not yet fully understood and known models are not able to reproduce the data. An alternative to a simulation is the use of real data events from the sidebands, i.e. the invariant mass regions close to the signal region, as training patterns for background events. The width of the signal region is limited by the detector resolution and can be deferred from a fit of the invariant mass spectrum (see section 5.2). Candidates taken from a sideband region where only a negligible number of signal events is expected provide a very pure source of combinatorial background events that are kinematically very similar to combinatorial background events in the signal region.
Ideally, an interpolation of the properties of candidates selected from the lower and upper sideband region would be desirable to closely reproduce the properties of combinatorial background events in the signal region. In the case of the neutral B mesons however, the lower sideband cannot be used to provide training patterns for background events, as the lower sideband generally contains contributions from partially reconstructed B meson decays, in addition to combinatorial background events. From figure 5.4 the contributions of several partially reconstructed B^0 decays in the final state $B^0 \to D\pi, D \to K\pi\pi$ can be identified: In the decay $B^0 \to D\rho, \rho \to \pi\pi^0$, the π^0

in the final state is not reconstructed, for $B^0 \to D^*\pi, D^* \to D\gamma$ and $D^* \to D\pi^0$, the γ or the π^0 is missing and the decays denoted $B^0 \to DX$ contain large contributions from the decays $B^0 \to De\nu$ and $B^0 \to D\mu\nu$ where the final state neutrino is not detected (see also section 5.2). These decays have, except for a lower invariant mass, very similar properties as fully reconstructed signal events.
Therefore events from the lower sideband in data cannot serve as patterns for combinatorial background events, instead only events from the upper sideband region in the invariant mass range from 5.4 to 5.6 GeV are used as training patterns for background events in the network training. Using events from the upper sideband only, it has to be made sure that these events are suited to describe the properties of combinatorial background in the signal region. The decays of B mesons into final states with four light particles (pions, kaons) are not limited by phase space, therefore events in the signal region and in the nearby upper sideband are kinematically very similar. In addition, variables used in the network training are selected avoiding a dependence on the invariant mass to make sure that the properties of combinatorial background events in the upper sideband region can be used to describe background events in the signal region (see also section 6.6.3).
Figure 6.1 visualizes the training samples selected from simulated events and data. For both samples soft preselection cuts are applied in order to reduce the size of the data sample by removing obvious background. In addition, a requirement on the minimum number of hits in the tracking devices is imposed to improve the quality of reconstructed candidates:

- Number of COT Axial/Stereo Hits $\geq$ 10/10,
- Number of Silicon Hits $\geq$ 3,
- all tracks have $p_t > 0.35$ GeV,
- $p_t(\pi_B) > 1.0$ GeV, $p_t(B) > 5.5$ GeV,
- $L_{xy}/\sigma_{L_{xy}}(D) > 6$, $L_{xy}/\sigma_{L_{xy}}(B) > 4$, $L_{xy}(B \to D) >$ -300 μm,
- $\chi^2_{r\phi}(D) < 20$, $\chi^2_{r\phi}(B) < 20$,
- $|d_0(B)| < 200\ \mu m$.

Figure 6.2 schematically summarizes the training and classification procedure. As first step the NeuroBayes® Teacher is used to perform the preprocessing of variables and the network training with input patterns for signal and background. Using the results of the training, the NeuroBayes® Expert calculates a network output for all events in the data sample obtained after the same preselection cuts. A possible bias resulting from using the same events in the upper sideband region for training and later after classification for the evaluation of the performance will be discussed in section 6.6.2.

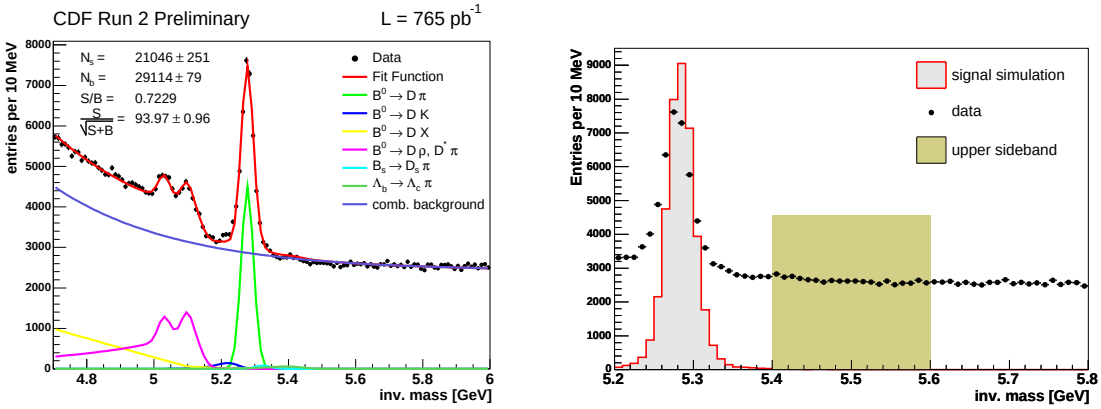


Figure 6.1: Composition of the training samples used as signal and background. Left: Composition of the input data sample after the preselection cuts are applied. Right: Schematical view of the data samples used for training, grey: simulated signal events, brown: upper sideband region providing training patterns for background events for the network training.

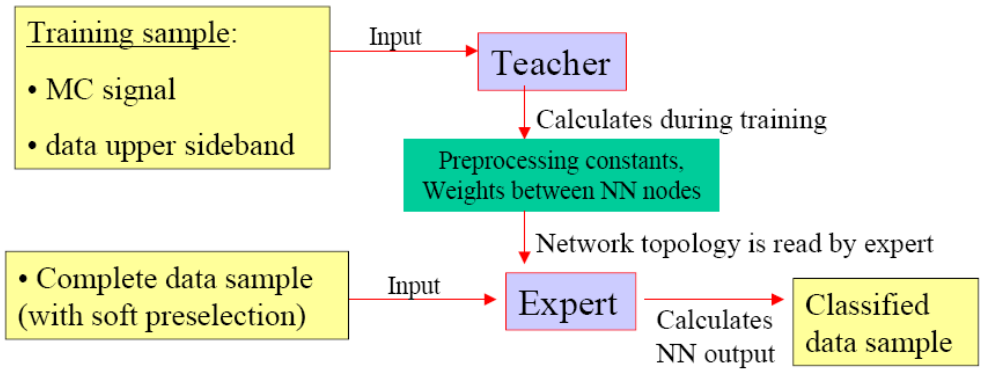


Figure 6.2: Procedure for the classification of a data sample. The Teacher uses the training samples to calculate the network topology and preprocessing constants that serve as input to the Expert for classification. In the end, a neural network output is obtained for each event in the data sample.

6.3 Obtaining Simulated Events

The sample of simulated B meson decays is obtained from $b\bar{b}$ events generated using the PYTHIA [14] event generator. $b\bar{b}$ creation at the Tevatron emerges via several production mechanisms (figure 6.3), that can be grouped into *flavor creation* (a,b,c), *flavor excitation* (d) and *gluon splitting* (e,f) processes. In order to simulate events at a center–of–mass energy of 1.96 TeV realistically, all of these processes have to be taken into account when generating $b\bar{b}$ pairs leading to the production of B mesons [61]. To produce the sample of simulated events that is used for the neural network training in the following sections, PYTHIA was run with the flag `msel=1` which enables all production mechanisms mentioned above. To produce a large sample of B meson decays in reasonable time, for each event the underlying event structure is kept and the b–hadrons are redecayed into selected final states using the `EvtGen` program package[62][57]. For the detector simulation `cdfSim`[63] is used, the output of `cdfSim` contains the same event structure and detector information as real data. The trigger decision is simulated with `TrigSim++`, where in particular `svtsim`[64] takes care of the simulation of the Two Track Trigger. The reconstruction of the simulated events proceeds in the same way as the reconstruction in real data (see section 4.2). After reconstruction and preselection cuts are applied to the sample of simulated events, 181k events of the decay $B^0 \to D\pi, D \to K\pi\pi$ and 41k events of the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ are available as signal patterns in the network training.

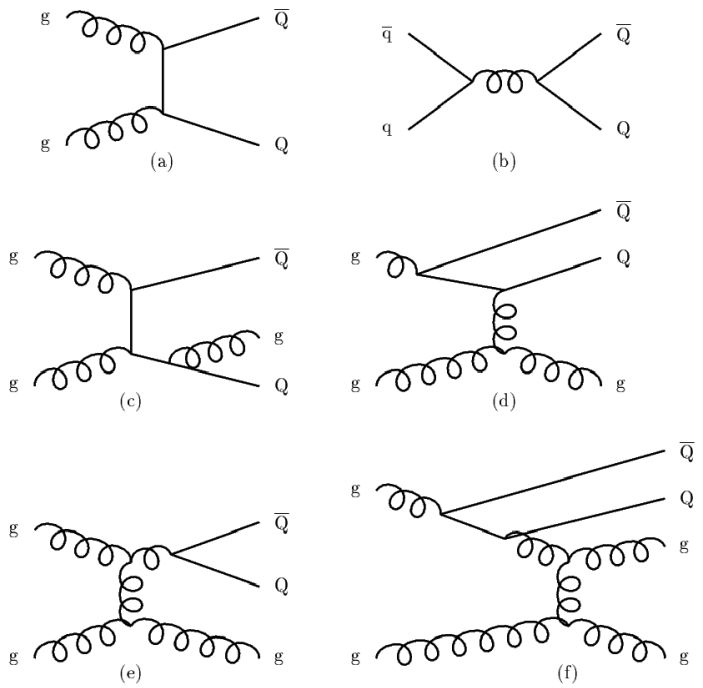
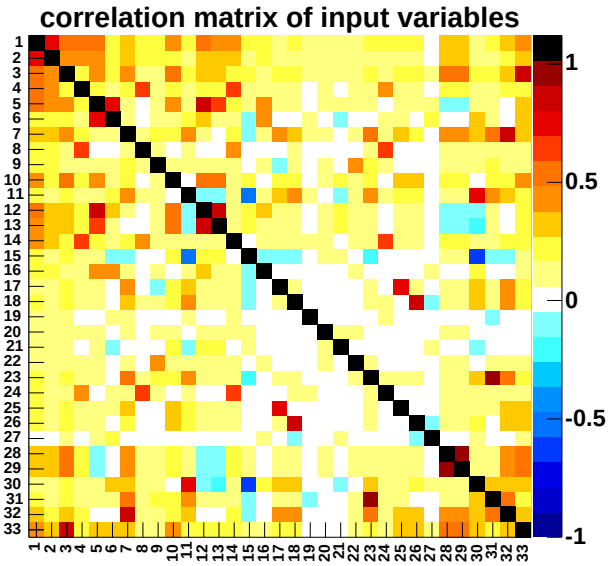


Figure 6.3: Examples of heavy-flavor production diagrams. (a,b) Leading order. (c) Pair creation (with gluon emission). (d) Flavor excitation. (e) Gluon splitting. (f) Events classified as gluon splitting but of flavor-excitation character (from [65]).

6.4 Selection of Training Variables

A large variety of variables is available which allows to discriminate between B meson signal events and combinatorial background. Many of these variables show a strong correlation and contain similar information. For example, the χ^2 of the B meson vertex fit is strongly correlated to the χ^2 of the D meson vertex fit, or the impact parameters of the four tracks involved in the decay show a strong correlation to the B mesons impact parameter $|d_0(B)|$ or its decay length $L_{xy}(B)$.
To determine the most important variables, a network training, including all variables that might contribute in some way to the signal optimization, is performed. The correlation matrix of the training variables is calculated as part of the NeuroBayes® preprocessing, each entry corresponds to the correlation coefficient $\rho_{ij} = \frac{cov_{ij}}{\sigma_i \sigma_j}$ obtained from the covariance matrix cov_{ij} of the variables. Figure 6.4 shows the correlation matrix of all training variables in a color coded representation. A detailed list providing definitions of all variables that are used can be found in appendix A. Many variables show strong correlations and by deploying sophisticated neural network based techniques, a significant improvement with respect to the cut based selection method can be expected.
To find out which and how many of these variables are actually necessary for an optimal performance of the network, a series of trainings is performed, removing successively the least significant variable from the list of variables given in table 6.1 until only the two most significant variables $|d_0(B)|$ and $L_{xy}/\sigma_{L_{xy}}(B)$ are used. The performance of a particular set of training variables is evaluated by the following procedure: Each training is used to calculate the network output for each event in the preselected data sample. The optimal network cut, maximizing $\frac{S}{\sqrt{S+B}}$ for each training, is obtained by applying sequentially harder cuts on the network output and fitting the resulting invariant mass distributions. The values for the number of signal events S and background events B in the region from 5.26 to 5.36 GeV around the central mass peak are derived from the fit and the quantities $\frac{S}{\sqrt{S+B}}$ and S/B are calculated.
Figure 6.5 summarizes the result, it shows the maximum signal significance $\frac{S}{\sqrt{S+B}}$ against the set of training variables used in the corresponding training. Apparently only few variables contribute significantly to the performance. Still these may be exchanged by others due to the redundance of the information provided by the training variables (see figure 6.4). Generally speaking, it's always better to limit the number of variables in the network training, as this improves the generalization capabilities (see section 3.2.2). In addition, there is the freedom to exclude certain variables from the network training due to a bad description in the simulation or due to an undesired dependence on the invariant mass of the B meson candidate (see also section 6.6.3).



Index	Name
1	target
2	$\lvert d_0(B)\rvert$
3	$L_{xy}/\sigma_{L_{xy}}(B)$
4	$\chi^2_{3D}(B)$
5	$\Delta R(D,\pi_B)$
6	$p_t(B)$
7	$\min \frac{\lvert d_0\rvert}{\sigma_{d0}}$
8	$\chi^2_{3D}(D)$
9	$\max(m_{K,\pi^1_D}, m_{K,\pi^2_D})$
10	$\lvert d_0(D)\rvert$
11	$L_{xy}/\sigma_{L_{xy}}(D)$
12	$p_t(\pi_B)$
13	$\sphericalangle(\vec{p}(B),\vec{p}(\pi_B))_{\mathrm{cms}B}$
14	$\chi^2_{r\phi}(B)$
15	$L_{xy}(B\to D)$
16	$\min p_t$
17	$\lvert d_0(K)\rvert$
18	$\lvert d_0(\pi^1_D)\rvert$
19	$p_t(\pi^2_D)$
20	$p_t(K)$
21	$p_t(D)$
22	$\min(m_{K,\pi^1_D}, m_{K,\pi^2_D})$
23	$\lvert d_0(\pi^2_D)\rvert$
24	$\chi^2_{r\phi}(D)$
25	lifetime–signed $d_0(K)$
26	lifetime–signed $d_0(\pi^1_D)$
27	$p_t(\pi^1_D)$
28	$\lvert d_0(\pi_B)\rvert$
29	lifetime–signed $d_0(\pi_B)$
30	$L_{xy}(D)$
31	lifetime–signed $d_0(\pi^2_D)$
32	$\min\lvert d_0\rvert$
33	$L_{xy}(B)$

Figure 6.4: Correlation matrix of the training variables. The entry *target* denotes the training target, in this case target$=1$ for signal and target$=-1$ for background events.

Rank	Name	Significance [σ]
1	$\lvert d_0(B)\rvert$	177.49
2	$L_{xy}/\sigma_{L_{xy}}(B)$	110.76
3	$\chi^2_{3D}(B)$	57.64
4	$\Delta R(D,\pi_B)$	47.01
5	$p_t(B)$	39.26
6	$\max(m_{K,\pi^1_D}, m_{K,\pi^2_D})$	19.76
7	$\min \frac{\lvert d_0\rvert}{\sigma_{d0}}$	16.93
8	$\lvert d_0(D)\rvert$	16.07
9	$p_t(\pi_B)$	15.94
10	$\chi^2_{3D}(D)$	12.59
11	$\sphericalangle(\vec{p}(B),\vec{p}(\pi_B))_{\mathrm{cms}B}$	9.97
12	$L_{xy}/\sigma_{L_{xy}}(D)$	7.45
13	$\lvert d_0(K)\rvert$	7.95
14	$\lvert d_0(\pi^1_D)\rvert$	9.09
15	$L_{xy}(B\to D)$	8.49
16	$\chi^2_{r\phi}(B)$	6.64
17	$p_t(D)$	6.38
18	$p_t(K)$	5.57
19	$p_t(\pi^2_D)$	5.67
20	$\min p_t$	5.04
21	$\min(m_{K,\pi^1_D}, m_{K,\pi^2_D})$	4.14
22	$\lvert d_0(\pi^2_D)\rvert$	3.20
23	lifetime–signed $d_0(K)$	2.97
24	$\chi^2_{r\phi}(D)$	2.85
25	$p_t(\pi^1_D)$	2.52
26	lifetime–signed $d_0(\pi_B)$	1.76
27	$\lvert d_0(\pi_B)\rvert$	1.81
28	$L_{xy}(D)$	0.72
29	lifetime–signed $d_0(\pi^2_D)$	0.67
30	lifetime–signed $d_0(\pi^1_D)$	0.55
31	$\min \lvert d_0\rvert$	0.42
32	$L_{xy}(B)$	0.43

Table 6.1: Loss of total correlation to the training target of all training variables. The numbers are calculated by successively taking out the variable that results in the smallest loss of total correlation to the target (see appendix B for more details).

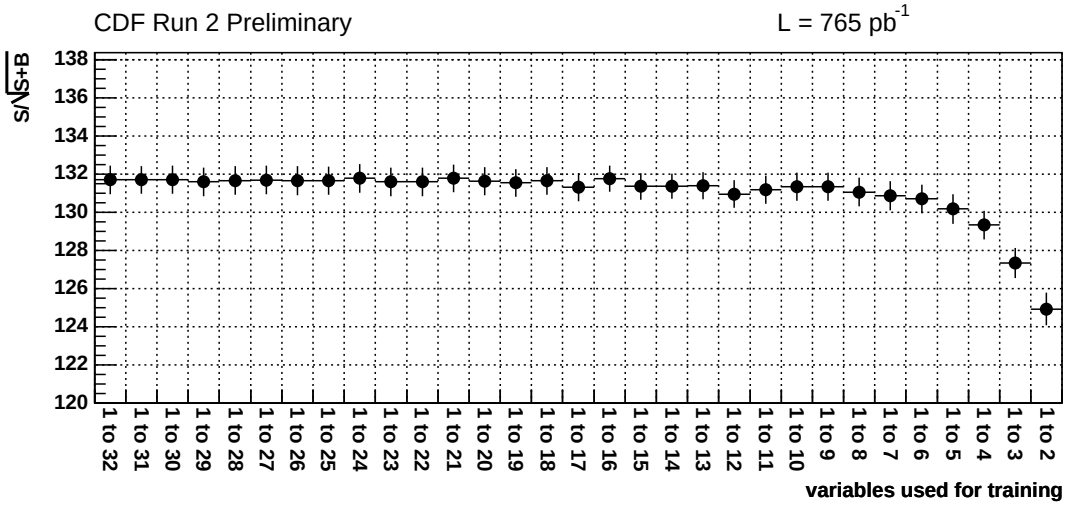


Figure 6.5: Performance of the network when removing the least significant variables one by one. The labels refer to the ranking of the variables as given in table 6.1.

The list of training variables given in table 6.1 includes several quantities that exhibit a strong dependence on the B meson's invariant mass *and* are not well described in the simulation. In detail, these variables are $p_t(K)$, $p_t(\pi_D^1)$, $p_t(\pi_D^2)$, $p_t(D)$, $p_t(B)$, $\Delta R(D, \pi_B)$ and $\sphericalangle(\vec{p}(B), \vec{p}(\pi_B))_{\mathrm{cms}B}$. These variables are removed from the network training. Comparisons of the distributions of all variables from sideband subtracted signal in data to simulated events can be found in appendix D. The mass dependence of the remaining variables is discussed in section 6.6.3.
The final set of variables is determined iteratively: Network trainings are performed, using the two most significant variables $|d_0(B)|$ and $L_{xy}/\sigma_{L_{xy}}(B)$ in combination with any of the remaining 23 training variables. The largest signal significances $\frac{S}{\sqrt{S+B}}$ obtained after the 23 trainings are compared. The variable resulting in the largest improvement for $\frac{S}{\sqrt{S+B}}$, when adding it to the set of two training variables, is kept for the next iteration. These three most important variables are then combined with each of the remaining 22 variables. This procedure is repeated until no further significant improvement by adding additional variables is observed. The set of ten variables obtained after this procedure are $|d_0(B)|$, $L_{xy}/\sigma_{L_{xy}}(B)$, $\chi^2_{3D}(B)$, $p_t(\pi_B)$, $\min \frac{|d_0|}{\sigma_{d0}}$, $L_{xy}/\sigma_{L_{xy}}(D)$, $L_{xy}(B \to D)$, $\chi^2_{r\phi}(B)$, $|d_0(D)|$ and $\min p_t$. Using these variables for the network training results in a significance of $\frac{S}{\sqrt{S+B}} = 131.5 \pm 0.9$. Adding further variables doesn't significantly improve $\frac{S}{\sqrt{S+B}}$. The significance obtained when using all 25 variables is 131.8 ± 0.8, which is consistent with the result obtained using the set of ten variables only.

6.5 Results for $B^0 \to D\pi, D \to K\pi\pi$

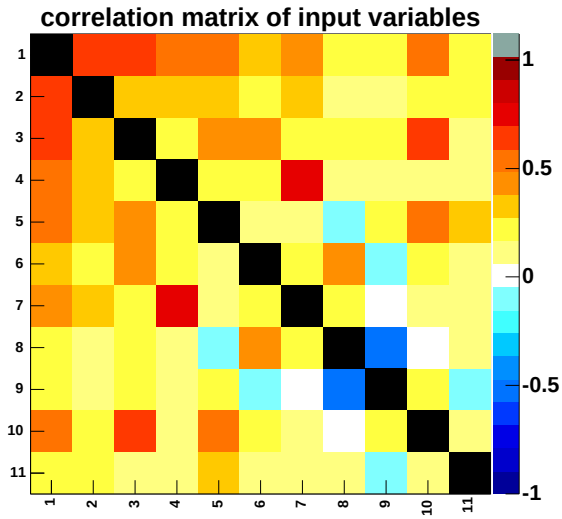
The previous section selected a set of ten variables for the optimization of the signal significance in the decay channel $B^0 \to D\pi, D \to K\pi\pi$. In this section, the results, that are obtained using this set of variables in the network training and the classification of the data sample are discussed in more detail. Also a comparison of the data samples obtained after the neural network classification and the cut based selection is performed.
Table 6.2 shows the loss of total correlation to the training target of the final set of ten training variables. Figure 6.6 shows the color coded representation of the correlation matrix of these variables, most the variables show some positive, in some cases rather strong, correlation, again underlining the fact that an improvement of the separation of signal and background can be expected from taking these correlations into account by using a neural network. Figure 6.7 shows two control plots created after the training of the network is completed, both demonstrate that the training was successful. Figure 6.8 compiles the number of signal events N_{sig}, the number of background events N_{BG}, the signal significance $\frac{S}{\sqrt{S+B}}$ and the ratio S/B derived from the fits of the invariant

mass spectra resulting from different cuts on the neural network output.
The curve displayed in the bottom left panel of figure 6.8 shows that it is possible to select a cut on the network output that yields a higher significance $\frac{S}{\sqrt{S+B}}$ than could have been obtained using a cut based selection. The maximum value is $\frac{S}{\sqrt{S+B}}$ =131.5 ± 0.9 found at a cut on the network output of $nn_{out} > 0.04$, the corresponding invariant mass distribution with fit is displayed in figure 6.9 (bottom). The improvement of the signal significance $\frac{S}{\sqrt{S+B}}$ amounts to $\simeq$ 11% [1] compared to $\frac{S}{\sqrt{S+B}}$ =118.3 ± 0.8 extracted from the fit of the invariant mass spectrum obtained after the cut based optimization procedure in figure 6.9 (top). Obtaining the same signal significance with the cut based optimization procedure would require 180 pb^{-1} or 24 % more data. For this estimate, the ratio of S/B obtained after the cut based optimization procedure is assumed to remain constant when more data is added.

Rank	Name	Significance [σ]
1	$\lvert d_0(B)\rvert$	177.49
2	$L_{xy}/\sigma_{L_{xy}}(B)$	110.76
3	$\chi^2_{3D}(B)$	57.64
4	$p_t(\pi_B)$	53.02
5	$\lvert d_0(D)\rvert$	19.37
6	$\min \frac{\lvert d_0\rvert}{\sigma_{d0}}$	18.06
7	$L_{xy}/\sigma_{L_{xy}}(D)$	8.33
8	$L_{xy}(B \to D)$	10.59
9	$\min p_t$	10.12
10	$\chi^2_{r\phi}(B)$	7.04

Table 6.2: Loss of total correlation to the training target of all training variables. The numbers are calculated by successively taking out the variable that results in the smallest loss of total correlation to the target (see appendix B for more details).

[1] The mass fit introduces an uncertainty in the order of 1% to the relative improvement.



Index	Name
1	target
2	$\lvert d_0(B)\rvert$
3	$L_{xy}/\sigma_{L_{xy}}(B)$
4	$\chi^2_{3D}(B)$
5	$p_t(\pi_B)$
6	$\min \frac{\lvert d_0\rvert}{\sigma_{d0}}$
7	$\chi^2_{r\phi}(B)$
8	$L_{xy}/\sigma_{L_{xy}}(D)$
9	$L_{xy}(B \to D)$
10	$\lvert d_0(D)\rvert$
11	$\min p_t$

Figure 6.6: Correlation matrix of the training variables. The entry *target* denotes the training target, in this case target$= 1$ for signal and target$= -1$ for background events.

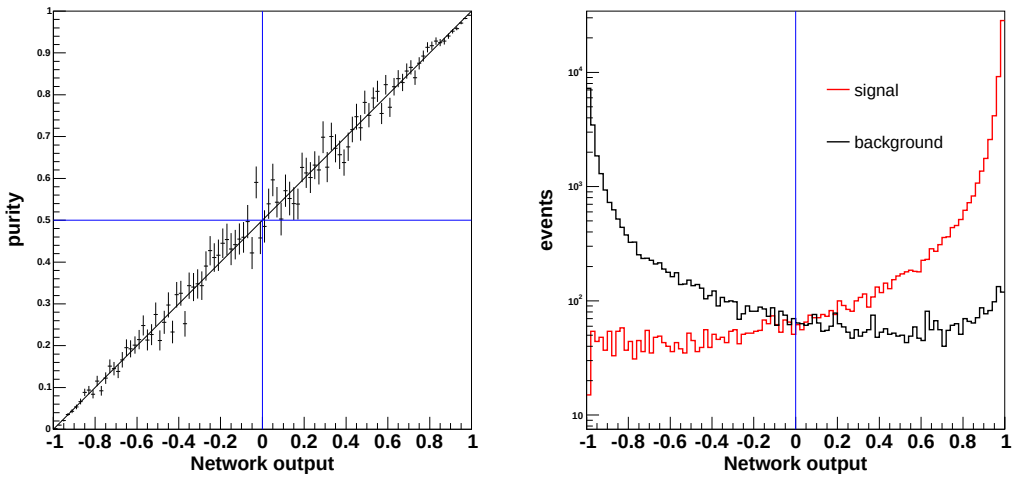


Figure 6.7: Quality plots from the NeuroBayes® training. The network output is a linear function of the network output (left) and signal and background training samples are well separated (right), two characteristics of a well–trained neural network [34].

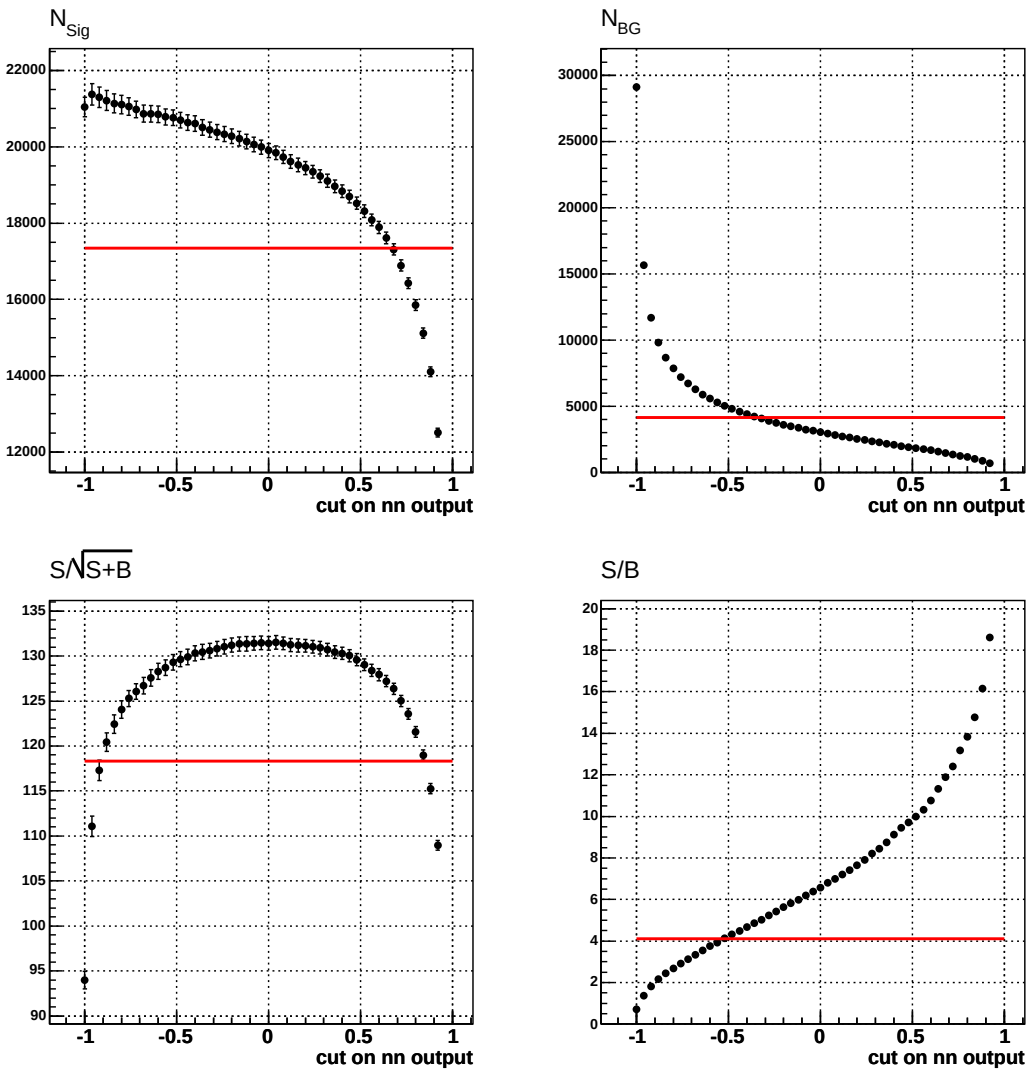


Figure 6.8: Number of signal events N_{sig}, number of background events N_{BG}, signal significance $\frac{S}{\sqrt{S+B}}$ and ratio of S/B derived from fits on the invariant mass spectra after applying different cuts on the neural network output. The red line in each panel marks the values of each quantity obtained from the cut based selection applied in the top panel of figure 6.9.

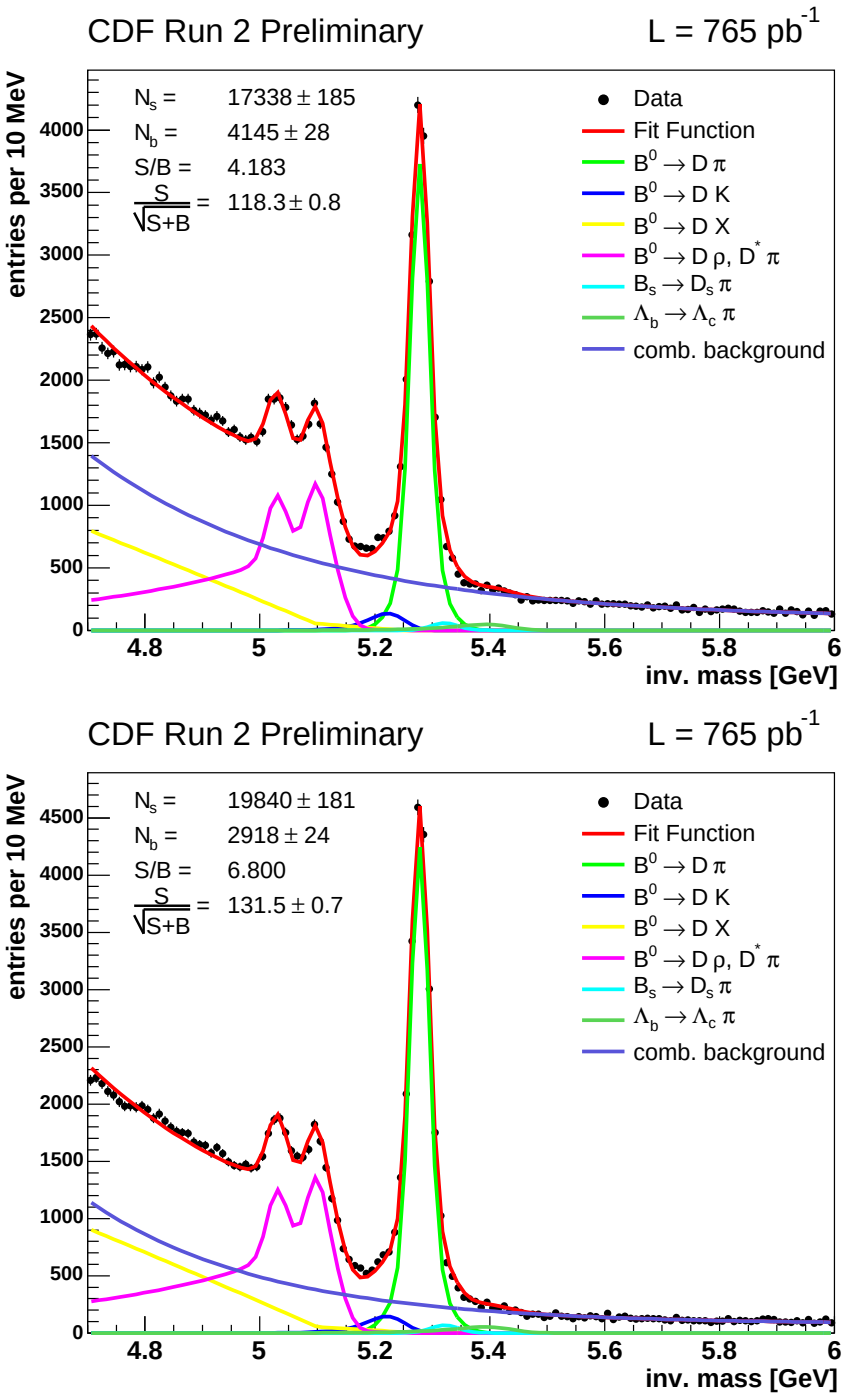


Figure 6.9: Invariant mass distribution obtained for the decay $B^0 \to D\pi, D \to K\pi\pi$ after applying the cuts from [49] (top) and applying a selection based on a neural network classification (bottom). In both cases the selection methods are optimized to obtain the maximum value for the signal significance $\frac{S}{\sqrt{S+B}}$.

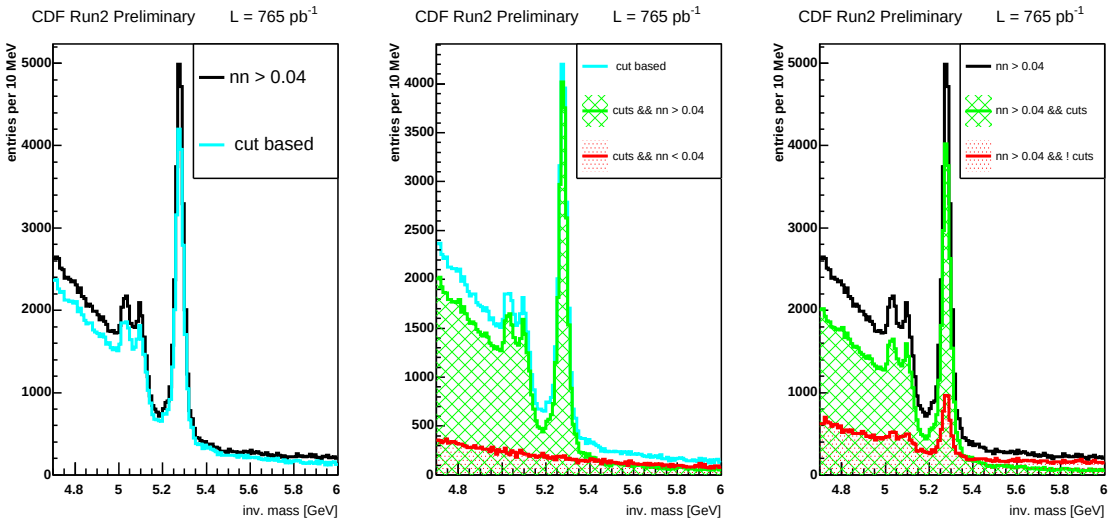


Figure 6.10: Direct comparison of events selected by sequential cuts and by the neural network. See text for a detailed explanation.

Finally, figure 6.10 compares the two samples obtained after a cut based selection and after a neural network based selection. The two distributions in the left panel correspond to the distributions already encountered in the top panel of figure 6.9 (blue, cut based) and the bottom panel of figure 6.9 (black, neural network). In both cases the selection methods are optimized to obtain the maximum value for the signal significance $\frac{S}{\sqrt{S+B}}$. The center panel shows two subsets of the sample obtained after cuts (blue): the green distribution contains events that both the neural network and the cuts would have selected, the distribution in red in the central panel denotes events that are part of the sample selected by cuts, but would have been rejected by the network cut. The shape of the red distribution indicates that these events are combinatorial background events mostly. In analogy, the distributions displayed in the right panel show the invariant mass distribution after applying the network cut (black) compared to the events that would have passed both cuts and network cut (green, same as in the center panel). The red distribution contains events that are lost when applying sequential cuts, but that are selected by the corresponding network cut. Here the shape of the red distribution indicates that the neural network method selects additional signal events that would have been rejected by the cut based selection.

6.6 Validation and Checks

This section summarizes miscellaneous checks that have been performed to understand and improve the neural network training procedure. First section 6.6.1 demonstrates that the particular choice of a sideband region doesn't play a significant role for the network performance. Section 6.6.2 shows that using events from the upper sideband in data for the network training and then again to evaluate the network performance after the classification of the data sample doesn't introduce any bias due to overtraining or memorizing training events. Possible effects originating in the dependence of the training variables on the candidates' invariant mass are investigated in section 6.6.3

6.6.1 Selection of the Sideband Region

Up to this point the selection of the sideband range from 5.4 to 5.6 GeV has been rather arbitrary. Due to kinematical differences, the distributions of the network training variables will have different shapes in different regions of the sideband, the sidebands closer to the signal region supposedly being more similar to the combinatorial background in the signal region.
To understand the effect of a particular choice of an upper sideband region, a series of checks has been performed. Figure 6.11 shows the dependence of the signal significance for events as training patterns for background from different 0.2 GeV wide slices of the sideband, starting with the range from 5.4 to 5.6 GeV up to range from 5.7 to 5.9 GeV. The particular choice of a sideband doesn't show any effect on the performance. Selecting the sideband region for the network training from 5.4 to 5.6 GeV is not a critical issue, other sideband regions close to the signal range will behave similarly in the training process and yield similar efficiencies when used for the signal selection. Figure 6.11 demonstrates that the use of different sideband regions doesn't bias the mass fit either, shown is the invariant mass of the B_s candidate obtained from the mass fit. Checks on other parameters don't indicate any bias either.

6.6.2 Bias Check

Using the same events from the upper sideband in data first as input patterns for background in the network training and then again, using this network training to decide if such an event is rather signal– or background–like, might introduce a bias. A possible source of such a bias can be an overtraining of the network. The effect of overtraining would result in a network that is not able to generalize from what it has learned but rather just memorize the training events. This would lead to suppressing background events that were part of the training process more rigorously than other background events (e.g. under the signal region) that have not been used in the

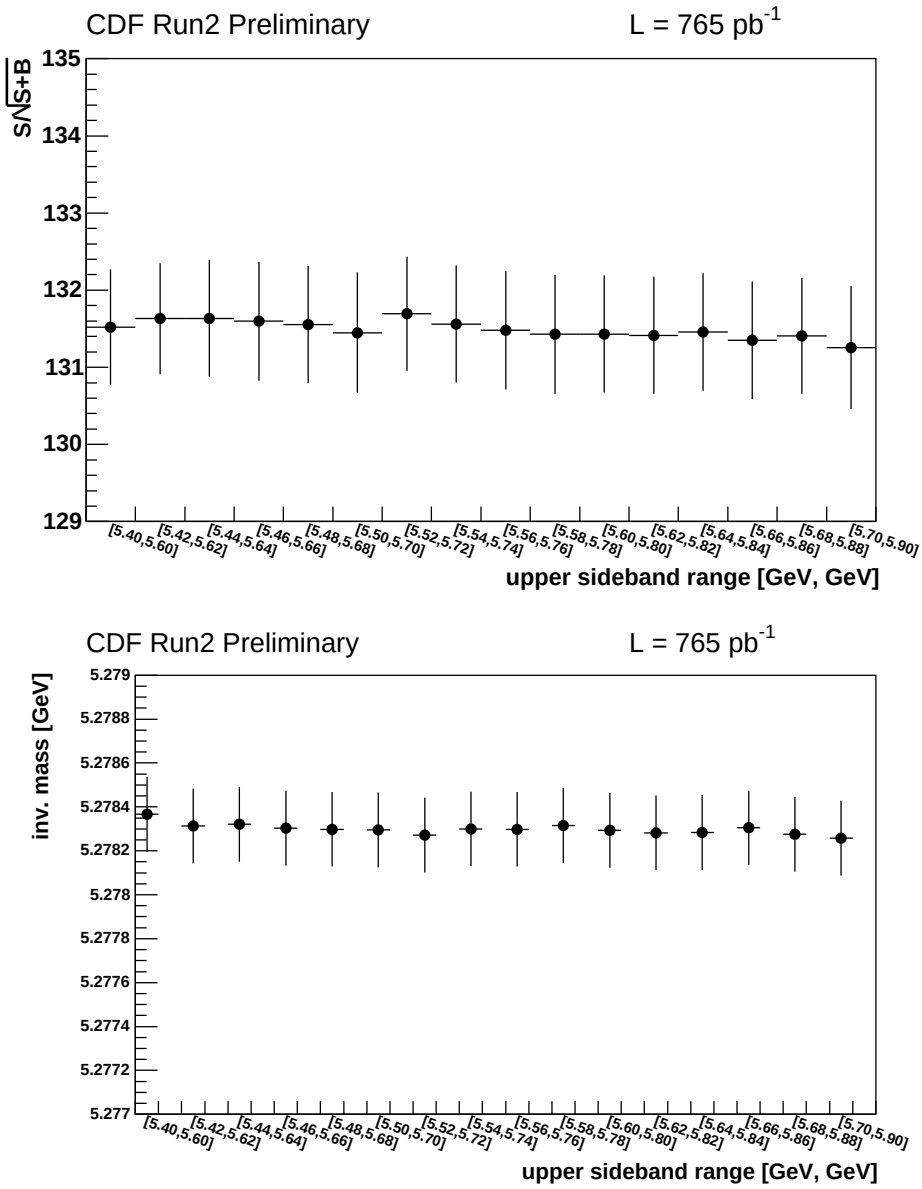


Figure 6.11: Maximum significance (top) and reconstructed B^0 mass (bottom) obtained for different slices of the upper sideband close to the signal region used for the network training.

network training. Regularization schemes prevent such an overtraining to occur when using NeuroBayes® (cf. section 3.2.3).
From figure 6.9 (bottom) it is already evident that a large effect induced by using identical events for training and classification can be excluded, otherwise the training region between 5.4 and 5.6 GeV would appear more suppressed than background from other sideband regions.
To evaluate this effect, the data sample is split into two statistically independent subsamples by selecting events with odd and even event numbers. These two subsamples are used as training patterns for background in two independent trainings, resulting in two independent sets of preprocessing constants and network weights, called the *Expertises* of NeuroBayes®. The Expertise from the training using the odd dataset is used to calculate network outputs for the even dataset and vice versa. Performing training and classification this way ensures that no event is classified from a network training where it has served as input pattern already. After the independent classification, the two classified samples are merged again and, after applying different network cuts, used as input for the fits of the invariant mass spectra. To compare the results obtained with this unbiased training/classification process with a dataset where a possible bias might occur from using identical events for training and classification, the Expertises from the even (odd) subsample are then used to classify the even (odd) subsample.
If the network had been over-trained and therefore memorized the patterns provided during the training process, it would show a tendency towards assigning smaller network outputs to the background events it already encountered during the training when identical subsamples are used for training and classification. For the sample prepared using different subsamples for training and classification this effect can't show up, as training and classification sample in this case are disjoint.
Figure 6.12 shows a comparison of the number of counted background events in the sideband region from 5.4 to 5.6 GeV, obtained after applying successively harder cuts on the network output on the two differently prepared data samples. No significant difference between the curves can be observed. Figure 6.13 shows two more exemplaric curves obtained in a similar way, denoting quantities derived from the fits of the invariant mass spectra after different network cuts. The top panel displays the number of signal events derived from the fits of the invariant mass spectrum for the two methods and the bottom panel compares the reconstructed B^0 masses. Again, no bias can be observed comparing the two methods.
Judging the results obtained from the tests presented in this section, using the same events from the upper sideband region in data for the network training that later is classified by the network as part of the full dataset. Figures 6.12 and 6.13 convince that the network does not memorize the events used in the network training as a result of overtraining.

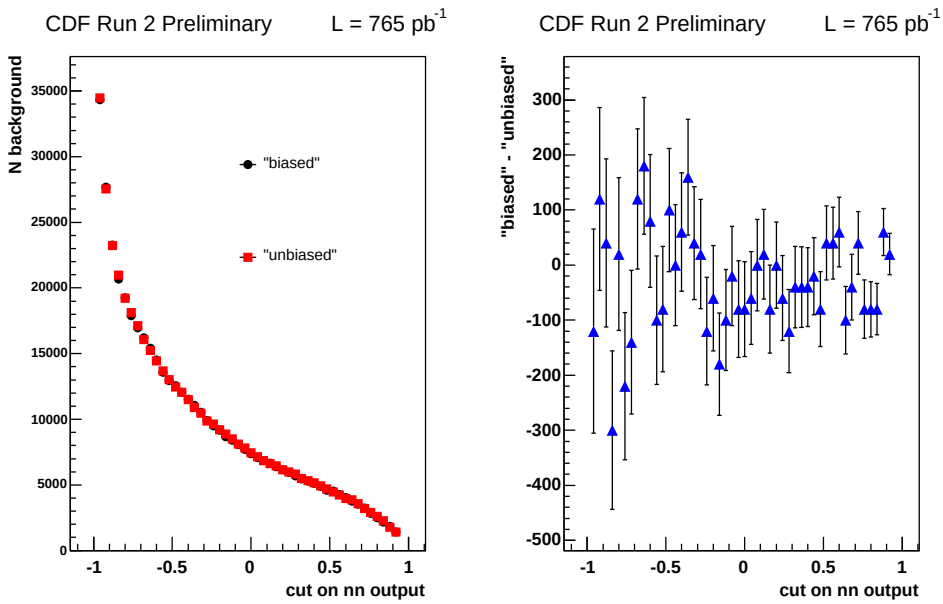


Figure 6.12: Left plot : Counted number of background events in the sideband region from 5.4 to 5.6 GeV used in the network training. The black circles denote the number of events N_{biased} counted when applying cuts on the neural network output for the sample that was obtained after using the same events as input for the network training and for classification of the sample ("biased"). The red squares show the number of events $N_{unbiased}$ in the training region obtained when using different events for training and classification ("unbiased"). Right plot: The difference of the two plots, the length of the error bars is equal to $\sqrt{N_{\text{biased}}}$.

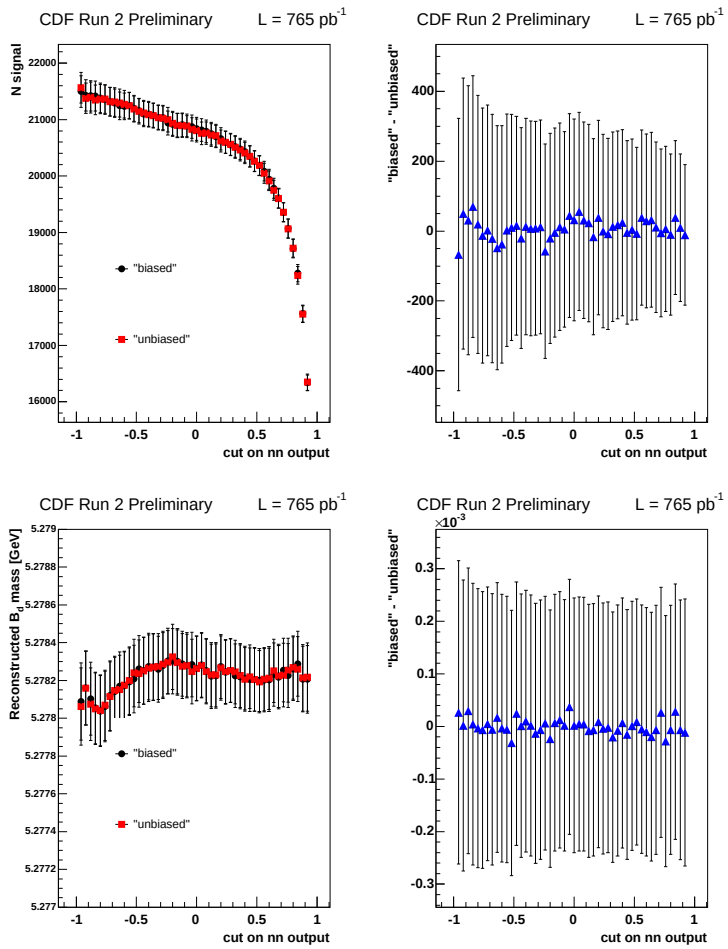


Figure 6.13: Top: Comparison of the number of signal events for the "biased" and "unbiased" training methods (left) and the difference of the two distributions (right). Bottom: Comparison of the fitted B^0 mass for the two training methods (left) and the difference of the two distributions (right).

6.6.3 Mass Dependence of Variables

Using a neural network trained with background events from the upper sideband in data and signal from simulated events implicates the possibility of the network not predominantly learning to distinguish B meson signal events from combinatorial background events, based on different physical properties of these events, but rather to learn that the difference between signal and background events only arises due to different invariant mass ranges of the training samples.
This has to be taken into account when selecting variables for training the network. In the worst possible case, the network would be able to fully reconstruct the invariant mass of a candidate. This would result in a suppression of all events outside the invariant mass region of the signal training sample and could be used to generate artificial signal peaks anywhere in the invariant mass spectrum by selecting events from this region as training patterns for signal and background events from a distinct invariant mass range.

Variables that exhibit a mass dependence

An example for a variable exhibiting a significant dependence on the invariant mass is the opening angle $\Delta R(D, \pi_B)$ between the D and π originating from a B meson decay. Figure 6.14 shows the mass dependence in a profile plot of the two–dimensional distribution of $\Delta R(D, \pi_B)$ against the candidates' invariant mass in the upper sideband. A similar mass dependence, though much less significant, can be seen for $L_{xy}/\sigma_{L_{xy}}(B)$ in the right panel of figure 6.14.
Figure 6.15 shows the behavior of the variable $p_t(\pi_B)$ for different sideband regions. The right panel compares the shapes of the $p_t(\pi_B)$ distributions in two different slices of the upper sideband. It can be seen that not only the mean values of the $p_t(\pi_B)$ distributions (left panel) show a dependence on the invariant mass but also the shape of the distributions vary. The trigger threshold observed at 2 GeV is less pronounced for higher invariant masses. Just rescaling the variable to flatten the means of the distributions will therefore not remedy this mass dependence. Many variables exhibiting such a mass dependence are excluded from the training (see section 6.4). The variable $p_t(\pi_B)$ however is used as training variable in the final set of variables (table 6.2), as it is an important variable contributing significantly to the performance of the network. Excluding $p_t(\pi_B)$ from the training, while keeping all other variables, reduces the maximum signal significance obtained from $\frac{S}{\sqrt{S+B}}$ =131.5 ± 0.9 to 130.2 ± 0.8[2]. Figure 6.16 demonstrates the effect introduced by the mass dependence of the training

[2]The values of $\frac{S}{\sqrt{S+B}}$, using or not using $p_t(\pi_B)$, are derived on the same events, so both values are strongly correlated. The difference of the performances is more significant than the uncertainty, derived from the fit of the invariant mass spectrum, suggests.

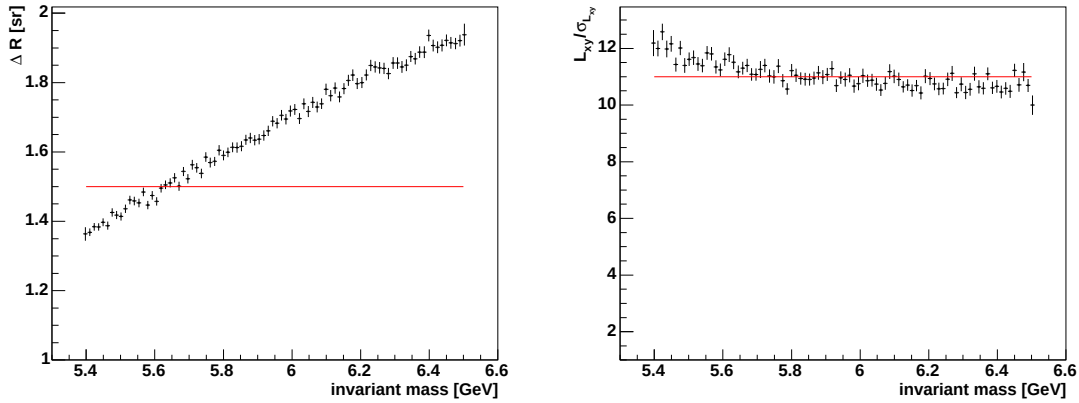


Figure 6.14: Profile plot of the two–dimensional distribution of $\Delta R(D, \pi_B)$ (left) and $L_{xy}/\sigma_{L_{xy}}(B)$ (right) against the candidates' invariant mass. A dependence on the invariant mass is observed. The red lines indicate the cuts on $\Delta R(D, \pi_B)$ and $L_{xy}/\sigma_{L_{xy}}(B)$ applied as part of the cut based optimization procedure.

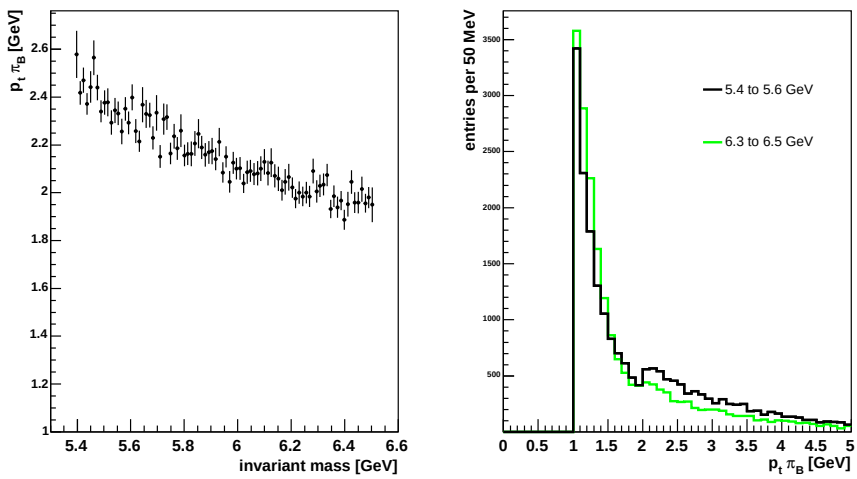


Figure 6.15: Left: Profile plot of the two–dimensional distribution of $p_t(\pi_B)$ against the candidates invariant mass. Right: $p_t(\pi_B)$ distributions for two sideband regions, it's evident that the shape of the $p_t(\pi_B)$ distribution changes within the upper sideband, the trigger threshold observed at 2 GeV becomes less pronounced when moving to higher invariant mass ranges.

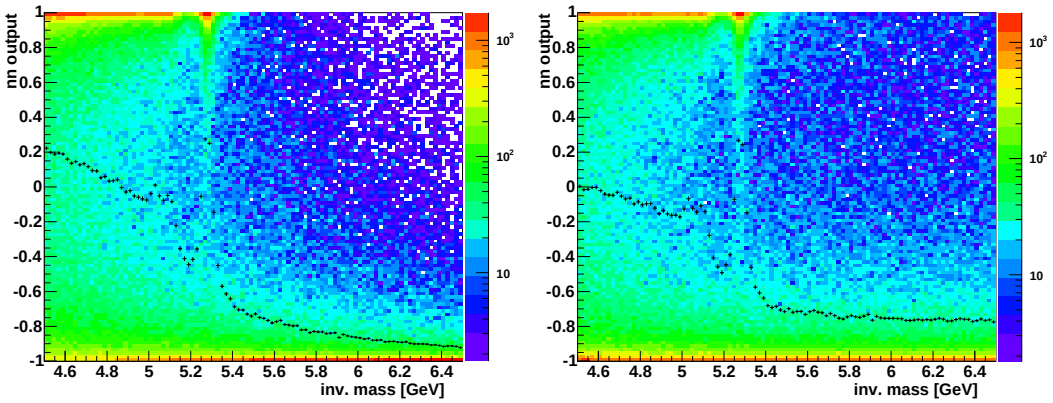


Figure 6.16: Distribution of the neural network output against the invariant mass of the B candidates. Left panel: Training including the full set of variables (table 6.1), including all variables with strong dependence on the invariant mass. Right panel: Training with the final set of variables (table 6.2). The black markers denote the mean value of the network output distribution in each invariant mass bin.

variables, if the full set of variables (table 6.1) is used instead of the final set of variables (table 6.2) in the network training. In the left plot (full set of variables), the network assigns higher network outputs to background events in the mass region closer to the signal region compared to background events with a higher invariant mass in the bottom right corner of the plot in the left panel. The plot in the right panel displays the distribution of network output against invariant mass for the final set of variables, excluding the significantly mass dependent variables $p_t(K)$, $p_t(\pi_D^1)$, $p_t(\pi_D^2)$, $p_t(D)$, $p_t(B)$, $\Delta R(D, \pi_B)$ and $\sphericalangle(\vec{p}(B), \vec{p}(\pi_B))_{\mathrm{cms}B}$ from the network training. The dependence of the network output on the invariant mass is reduced, though still some preference for lower invariant masses is seen.

Training with two sidebands

To determine which variables out of the final set of ten variables used in section 6.5 show a correlation to the invariant mass, two sideband regions are used as input patterns for signal and background in a network training. To mimic the effect of using a signal region from roughly 5.2 to 5.4 GeV and a background region in the range from 5.4 to 5.6 GeV, a training with two similarly wide slices of the upper sideband is performed. Sideband events from 5.5 to 5.7 GeV are fed as signal to the network and events from the range 5.7 to 5.9 GeV are declared to be background in the training.

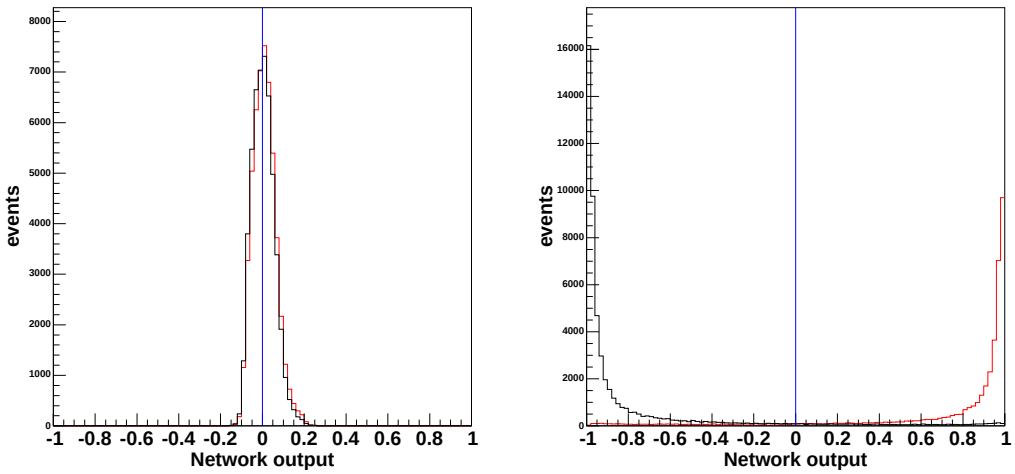


Figure 6.17: Left: Network output for the training sample composed of the sideband region from 5.5 to 5.7 GeV fed in as signal (red) and the region from 5.7 to 5.9 GeV declared as background (black). Right: Network outputs obtained after a training using simulated B meson decays as signal (red) and events from the upper sideband region from 5.4 to 5.6 GeV as background (black).

Using these ranges will show if the network is able to produce an artificial signal peak in the sideband region used as signal in the network training.
The left plot in figure 6.17 shows the network outputs obtained for the two training samples in the upper sideband. Ideally, if none of the training variables showed any mass dependence, the two distributions would lie on top of each other. As can be seen, the network has learned something, as the red curve for signal from the sideband region 5.5 to 5.7 GeV is slightly shifted towards higher network outputs. Comparing this separation to the separation of the distributions displayed in the right panel, where B meson decays from a simulation and events in the upper sideband from 5.4 to 5.6 GeV are used as training samples, this is a very small effect. Table 6.3 shows the loss of total correlation to the training target of the variables obtained after using the two slices of the upper sideband as training input. As expected the variable $p_t(\pi_B)$ shows the capability to separate the two sideband because of the modified shape of the distribution when moving to higher invariant masses.
To demonstrate the impact of this mass dependence on the invariant mass distribution, the full upper sideband region from 5.4 to 6.5 GeV is classified using the training obtained with the two sideband slices used as training samples. The plots in figure 6.18 demonstrate the behavior of the classified upper sideband distribution by applying

Rank	Name	Significance [σ]
1	$p_t(\pi_B)$	12.61
2	$\lvert d_0(D)\rvert$	4.49
3	$\lvert d_0(B)\rvert$	4.27
4	$L_{xy}/\sigma_{L_{xy}}(B)$	4.47
5	$\min \frac{\lvert d_0\rvert}{\sigma_{d0}}$	3.23
6	$L_{xy}/\sigma_{L_{xy}}(D)$	1.73
7	$L_{xy}(B \to D)$	0.68
8	$\chi^2_{3D}(B)$	0.25
9	$\chi^2_{r\phi}(B)$	0.08
10	$\min p_t$	0.03

Table 6.3: Loss of total correlation to the target of variables when using two different slices of the upper sideband as training samples for signal and background.

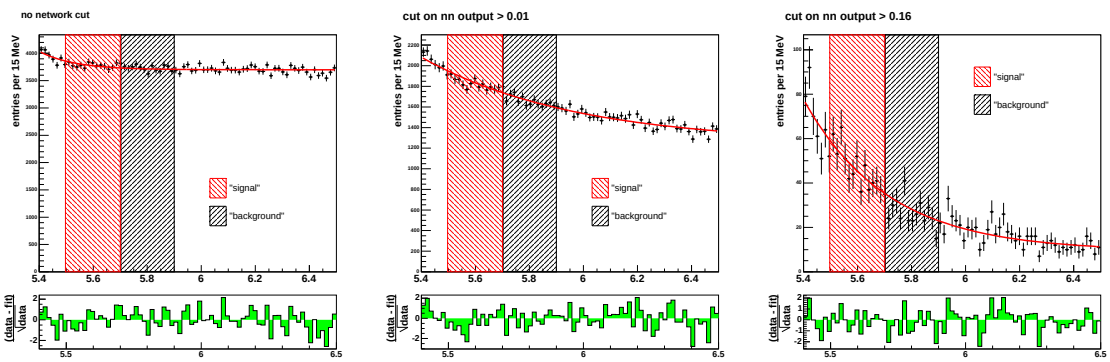


Figure 6.18: Invariant mass distributions obtained when using two slices of the upper sideband are used in the network training and the full upper sideband sample is classified using this training. The boxes denote the sideband regions fed as "signal" and "background" to the network training. Left: inclusive distribution, no network cut applied. Center and right panel: mass distributions after applying a weak and a hard cut on the network output. The function fitted to the distributions corresponds to $P_{comb}(m)$, the parametrization that is chosen to describe combinatorial background events in the fit of the full invariant mass spectrum (see section 5.2.1).

cuts on the neural network output. The training region from 5.5 to 5.7 GeV is not emphasized compared to the regions either from 5.7 to 5.9 GeV, which was used as background in the training process, or from 5.4 to 5.5 GeV, that was not part of the training sample. From figure 6.18 one can conclude that the network is not capable of producing spurious signal peaks simply by selecting different sideband regions as training patterns for signal and background.

Comparing neural network and cuts

Another effect visible in figure 6.18 is the preference of the neural network for lower invariant masses, when trained with two slices of the upper sideband as samples of signal and background events. The remaining mass dependence of the training variables in the network training is not be capable of producing a spuriously enhanced signal peak, but rather modify the shape of the background spectrum with a preference for lower invariant masses. This modified shape is taken into account by the parameters of the function used to describe the distribution of combinatorial background events as part of the fit of the invariant mass spectrum (cf. eq. 5.9). Using $\Delta R(D, \pi_B)$, $p_t(B)$ and $p_t(\pi_B)$ as ingredients of the cut based optimization procedure, such a bias towards lower invariant masses must have been inherent to all previously published analyses already (this is visualized by the red lines in figure 6.14, indicating cuts that are applied as part of the cut based optimization procedure). Figure 6.19 compares the shapes of the background distributions obtained after a cut based selection and after the neural network selection.
Beyond a slight modification the shape of the background distribution, no further effect of the mass dependence of some of the training variables can be observed. The shape of the background distribution shows a similar behavior as already observed during the cut based optimization procedure, therefore no special treatment of variables that exhibit some dependence on the invariant mass is considered.

6.7 Results for $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$

The optimization strategy developed in previous sections for the signal significance of the decay $B^0 \to D\pi, D \to K\pi\pi$ can be applied to other B meson decays as well. The necessary ingredients are a simulation of the resonance decay, providing training patterns for signal events, and combinatorial background events from the upper sideband in data, serving as patterns for background events in the network training. Of course, each decay channel exhibits specific features that can be taken into account when selecting variables for the network training. For example, the angular correlations between daughter and parent particles in the decay chain or resonant substructures of

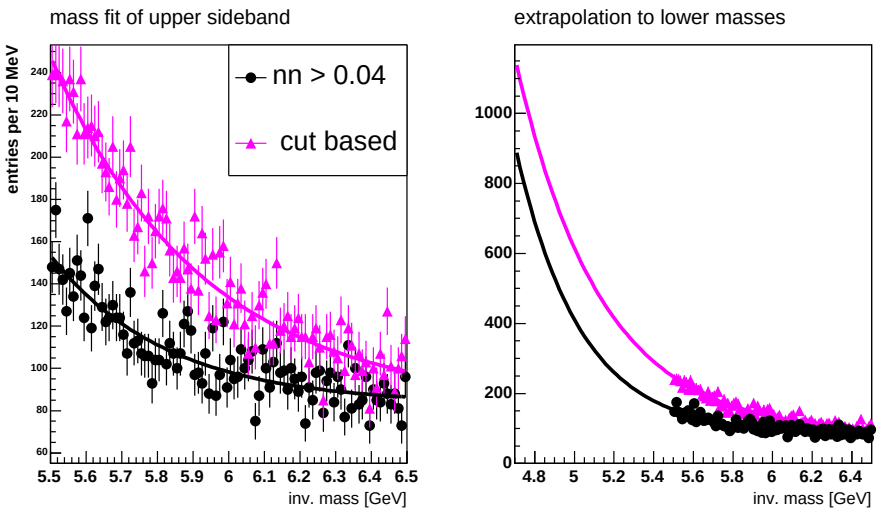


Figure 6.19: Comparison of the distributions of invariant mass in the upper sideband obtained after the cut based optimization procedure (magenta triangles) and a neural network selection (black circles). In both cases, the distributions rise with lower invariant masses, a fit on the background distributions shows that the shape of both curves is similar. Left: mass range from 5.5 to 6.5 GeV where the fit is performed, right: extrapolation to lower invariant masses.

the daughter particles can contribute additional information to the network training in different decay modes.
This section focuses on the decay of the B_s meson into $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$. As the optimization procedure developed in this work is intended to improve the existing Δm_s measurement at CDF II, it will be demonstrated that the use of a neural network for the already intensively studied decay mode (see e.g. [60][58]) can still significantly improve the signal significance $\frac{S}{\sqrt{S+B}}$ relevant for a measurement of Δm_s. To reduce the size of the data sample and to improve the track quality by imposing requirements on the number of detector hits, a set of soft preselection cuts is applied:

- Number of COT Axial/Stereo Hits $\geq 10/10$,
- Number of Silicon Hits ≥ 3,
- all tracks have $p_t > 0.35$ GeV,
- $p_t(B) > 5.5$ GeV,
- $L_{xy}/\sigma_{L_{xy}}(D) > 2$, $L_{xy}/\sigma_{L_{xy}}(B) > 2$
- $\chi^2_{r\phi}(D) < 20$, $\chi^2_{r\phi}(B) < 20$,
- $|d_0(B)| < 200\ \mu m$,
- $p_t(\pi_B) > 1.0$ GeV.

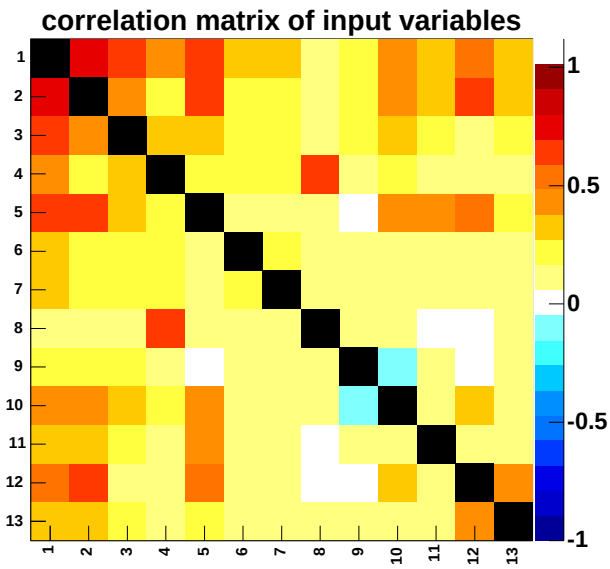
The set of training variables used to achieve an optimal performance for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ is obtained after an iterative procedure similar to the method applied for the B^0 meson in section 6.4. Variables are added to the network training step by step until no further improvement can be observed. Table 6.4 displays the set of variables obtained by this procedure for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$. The set of variables table 6.4 is very similar to the set of variables obtained for the decay mode $B^0 \to D\pi$ in table 6.2 in section 6.5. This reflects the fact that in both cases, the decay of long–lived B mesons is separated from combinatorial background events by similar properties. On the other hand, table 6.4 also demonstrates that in the case of the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ some additional variables can contribute significantly which don't contribute in the case of the decay $B^0 \to D\pi$. These additional variables are the helicity angle of either of the kaons from the decay $\phi \to KK$ and the invariant mass of the ϕ, which reflects the different angular correlation and subresonance structure of the daughter particles of the decay $D_s \to \phi\pi$ compared to $D \to K\pi\pi$. Figure 6.20 shows the correlation of the training variables and figure 6.21 displays the NeuroBayes® training quality plots. Signal and background are well separated in the training sample and the network output is a linear

function of the purity, indicating that the training was successful. Training patterns for background used in the network training are taken from the upper sideband region from 5.6 to 5.9 GeV.

Rank	Name	Significance [σ]
1	$L_{xy}/\sigma_{L_{xy}}(B)$	159.21
2	$\lvert d_0(B)\rvert$	62.46
3	$p_t(\pi_B)$	30.72
4	$\chi^2_{3D}(B)$	26.09
5	helicity angle K^1	14.51
6	$L_{xy}(B \to D)$	13.85
7	$m_{\phi\to KK}$	12.48
8	$\lvert d_0(D)\rvert$	11.97
9	$\min p_t$	11.54
10	$\chi^2_{3D}(D)$	10.11
11	$L_{xy}/\sigma_{L_{xy}}(D)$	9.27
12	lts. $d_0(K^1)/\sigma_{d0}$	3.16

Table 6.4: Training variables used for the optimization of the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$.

Figure 6.22 shows the number of signal events N_{sig}, the number of background events N_{BG}, signal significance $\frac{S}{\sqrt{S+B}}$ and ratio of S/B derived from fits on the invariant mass spectra after applying different cuts on the neural network output. Figure 6.23 (bottom) shows the mass distribution for a network cut of 0.76, this yields the largest value for $\frac{S}{\sqrt{S+B}}$ of 35.5 ± 0.7. This corresponds to an improvement of $\simeq 16\%$ compared to the cut based optimization procedure [60] (displayed in figure 6.23, top panel) that results in $\frac{S}{\sqrt{S+B}} = 30.6 \pm 0.8$. This improvement is equivalent to adding 260 pb^{-1} or 35 % more data, when using the cut based optimization procedure instead of the neural network selection to obtain the same performance. The expression describing the overall signal significance of a Δm_s–measurement in eq. 5.31 not only depends on the absolute number of signal and background events, contained in the factor $\frac{S}{\sqrt{S+B}}$. For a time–resolved measurement of B_s oscillations, in particular the proper decay time resolution σ_t plays an important part in determining the overall signal significance (see section 5.6). To estimate the proper time resolution attenuation factor obtained for signal events in the decay channel $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ after applying a



Index	Name
1	target
2	$L_{xy}/\sigma_{L_{xy}}(B)$
3	$\lvert d_0(B)\rvert$
4	$\chi^2_{3D}(B)$
5	$p_t(\pi_B)$
6	helicity angle K^1
7	$m_{\phi\to KK}$
8	$\chi^2_{3D}(D)$
9	$L_{xy}/\sigma_{L_{xy}}(D)$
10	$L_{xy}(B \to D)$
11	$\min p_t$
12	$\lvert d_0(D)\rvert$
13	lts. $d_0(K^1)/\sigma_{d0}$

Figure 6.20: Correlation matrix of the training variables for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$. The entry *target* denotes the training target, in this case target$= 1$ for signal and target$= -1$ for background events in the training.

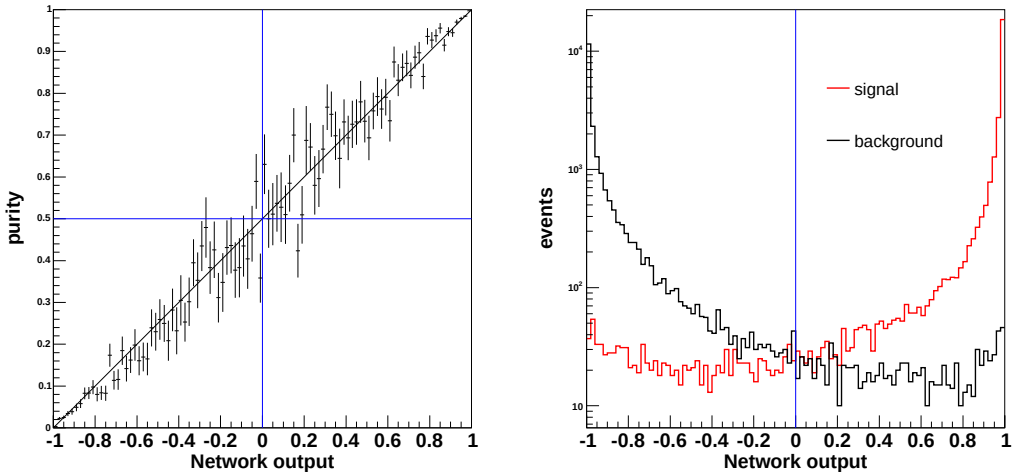


Figure 6.21: Quality plots from the NeuroBayes® training for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$. The purity is a linear function of the network output (left) and signal and background training samples are well separated (right), two important characteristics of a well–trained neural network.

particular selection procedure, the weighted average

$$\mathcal{S}_t = \frac{1}{N}\sum_{i=0}^{N} e^{-\frac{1}{2}\sigma_{t,i}^2\,\Delta m_s^2} \tag{6.1}$$

is calculated as the sum over all simulated events i passing the selections. For Δm_s a value of 17 ps^{-1} is assumed. Figure 6.24 (left) compares the values of $\mathcal{S}_t$ obtained after different neural network selection cuts and a cut based optimization procedure. The maximum value for $\mathcal{S}_t \times \frac{S}{\sqrt{S+B}}$ from fig. 6.24 corresponds to an improvement of $\simeq 14\%$, which is slightly less than the $\simeq 16\%$ obtained when taking into account $\frac{S}{\sqrt{S+B}}$ only.

6.8 Summary

Using a neural network trained with simulated signal events and events from the upper sideband in data can increase the signal significance $\frac{S}{\sqrt{S+B}}$ for exclusively reconstructed final states of B meson decays significantly. The method can be applied to all exclusive decay modes. The performance of the network selection can be optimized by taking into account individual characteristics of each decay channel. A wide

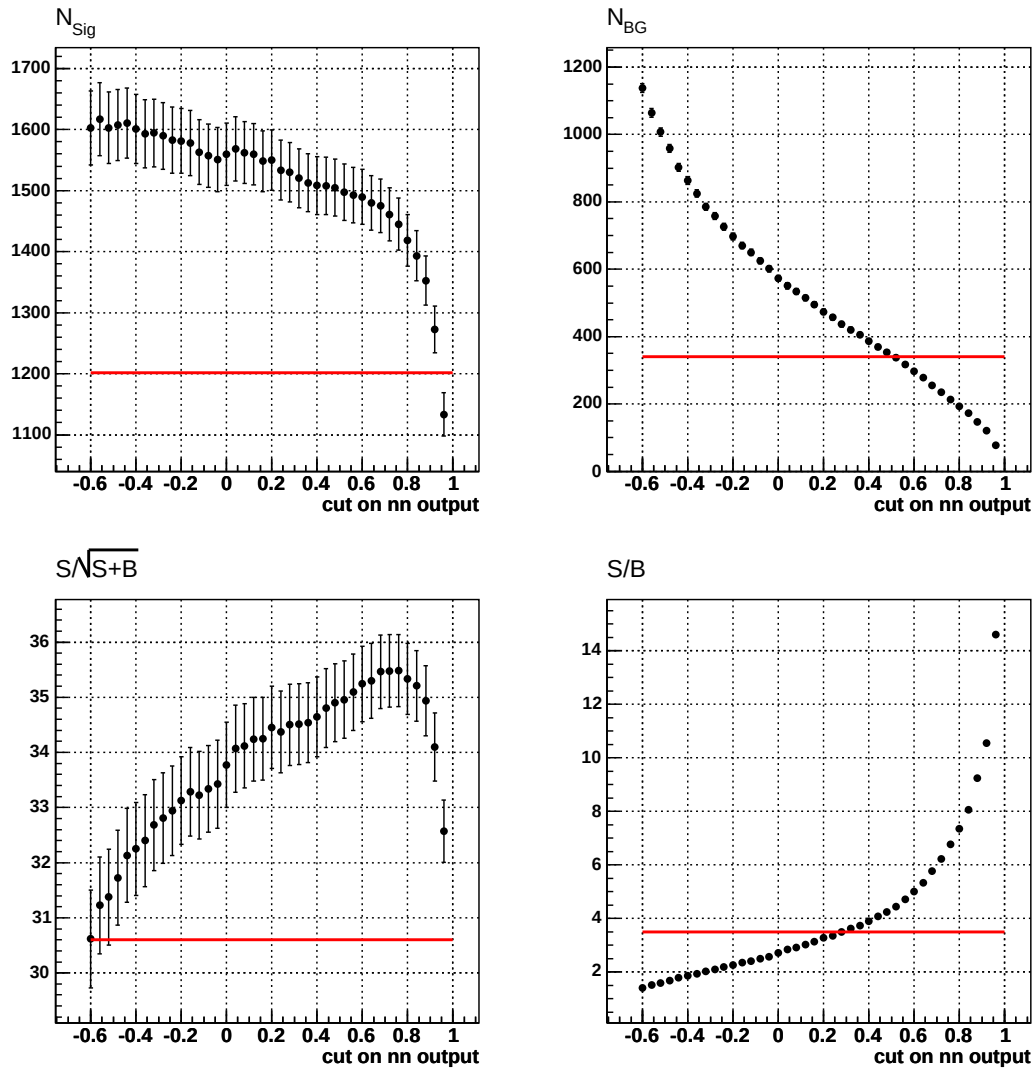


Figure 6.22: Number of signal events N_{sig}, number of background events N_{BG}, signal significance $\frac{S}{\sqrt{S+B}}$ and ratio of S/B derived from fits on the invariant mass spectra after applying different cuts on the neural network output. The red line in each panel marks the values of each quantity obtained from the cut based selection applied in the top panel of figure 6.23 (top).

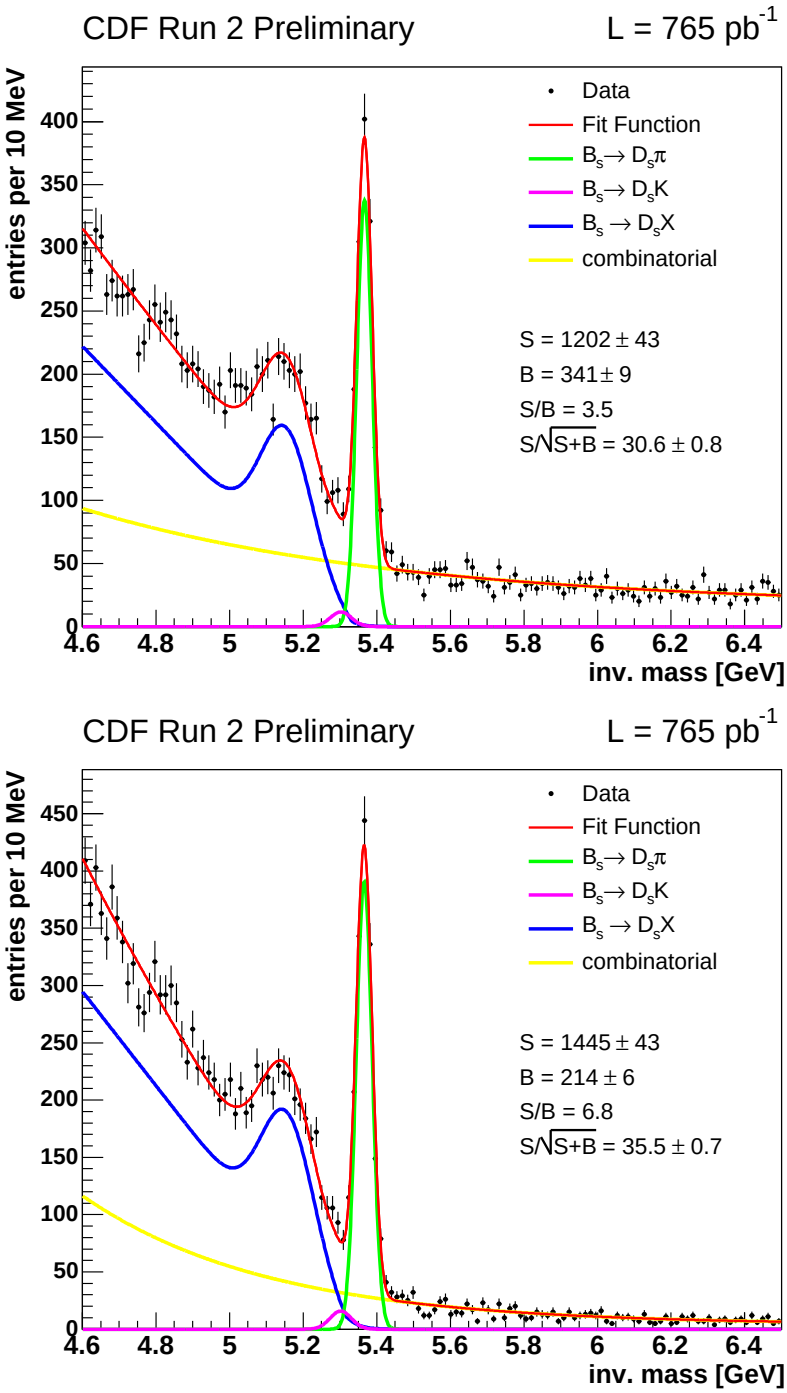


Figure 6.23: Invariant mass distribution obtained for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ after applying the sequential cuts from [60] (top) and using a selection based on a neural network classification (bottom). In both cases the selection methods are optimized to obtain the maximum value for the signal significance $\frac{S}{\sqrt{S+B}}$.

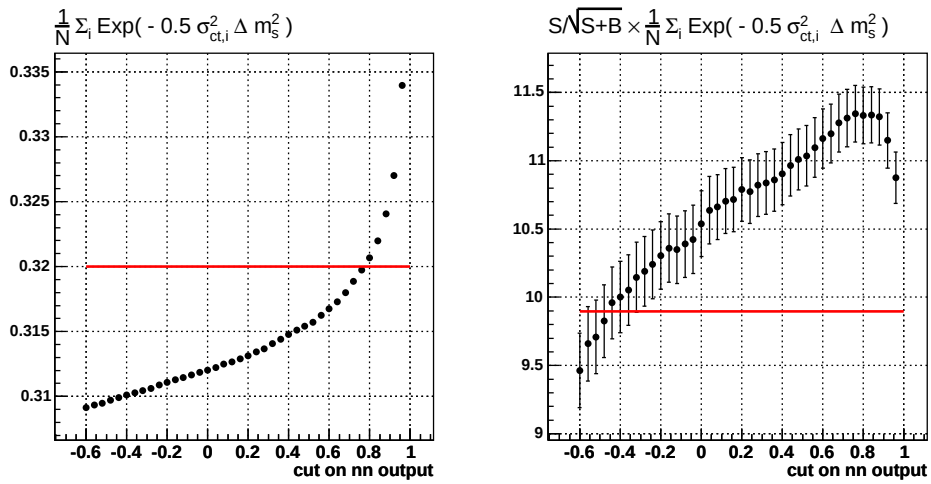


Figure 6.24: Left panel: Comparison of the contribution $\mathcal{S}_t$ of the proper time resolution σ_t estimated from simulated events of the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ for different neural network selection cuts (black markers) and using the cut based selection (red line). Right panel: $\mathcal{S}_t \times \frac{S}{\sqrt{S+B}}$ for different neural network selection cuts (black markers) and the cut based selection (red line).

range of systematic effects were evaluated for the decay mode $B^0 \to D\pi, D \to K\pi\pi$. It has been demonstrated that neither a weak dependence on the invariant mass of some training variables, nor the specific choice of a sideband region nor the overlap between training and classification sample introduces any systematic bias in the data sample. The signal significance $\frac{S}{\sqrt{S+B}}$ for the decay mode $B^0 \to D\pi, D \to K\pi\pi$ could be improved by 11 %, which is equivalent to adding 180 pb^{-1} or 24 % more data to the existing 765 pb^{-1} when the cut based optimization procedure is used to obtain the same performance.
For the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$, the significance $\frac{S}{\sqrt{S+B}}$ could be improved by 16 %. This improvement is equivalent to adding 260 pb^{-1} or 35 % more data, when using the cut based optimization procedure instead of the neural network selection to obtain the same performance. Taking into account the proper decay time resolution σ_t of the selected events, an overall improvement of 14 % in the significance of the measurement of Δm_s in this decay channel can be expected.
Beyond an application in a B^0 or B_s mixing measurement, the selection method developed in this chapter can significantly enhance many physics analyses using exclusively reconstructed B mesons. Depending on the particular requirements of each analysis, an optimal working point can be determined choosing the neural network selection cut appropriately. The curves in the bottom right panels of the figures 6.8 and 6.22

demonstrate that, for example, the signal purity, expressed as S/B, can be made almost arbitrarily large by applying very hard cuts on the neural network output.

Chapter 7

Mixing Fit

7.1 Overview

The fit framework described in chapter 5 is applied to the data samples obtained after performing different types of selections in this section and used to derive a value for Δm_d (Δm_s) from the sample containing the decay mode $B^0 \to D\pi, D \to K\pi\pi$ ($B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$). The general outline of the fit procedure for the decay mode $B^0 \to D\pi$ can be summarized as follows:

1. Apply all selection cuts to the data sample, simulated signal events and simulated background components (partially reconstructed B mesons and final states with misassigned particles).

2. Perform a binned maximum likelihood fit of the full invariant mass spectrum:

   (a) Fit of the mass template functions for partially reconstructed B mesons and final states with misassigned particle hypothesis.

   (b) Full mass fit of the invariant mass spectrum.

3. Perform an unbinned maximum likelihood fit of the B meson lifetime:

   (a) Fit of the corresponding lifetime template functions for final states with misassigned particle hypothesis.

   (b) Fit of the lifetime efficiency curve.

   (c) Perform an unbinned maximum likelihood fit in mass and lifetime space simultaneously, both spaces share the background fraction f_B as common parameter.

4. Perform an unbinned maximum likelihood fit of Δm_d (Δm_s).

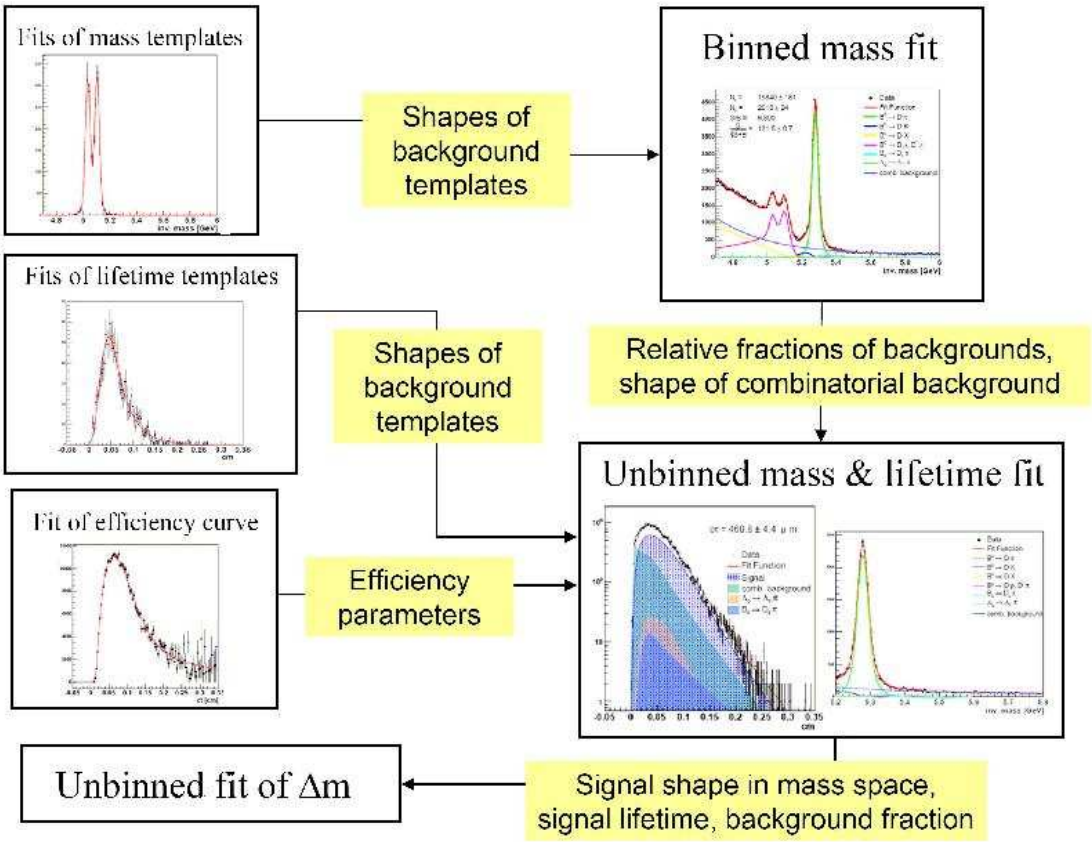


Figure 7.1: Flow chart summarizing the fit procedure. The yellow boxes, superimposed on the arrows, display the fit parameters that are fixed after each step.

Figure 7.1 shows a flow chart displaying the sequence of steps of the fit procedure. In the following sections, a selection of events based on the cut based optimization procedure and a neural network selection, as outlined in chapter 6, are used to select the input data sample used for the fit. It will be demonstrated that the mixing fit framework can be used with both selections to extract values for the B^0 meson lifetime and Δm_d that are consistent with the world average. Section 7.4 will apply a similar fit procedure to compare the cut based and the neural network based selection methods by using events selected by either method as input for an amplitude scan and an unbinned maximum likelihood fit of Δm_s. The mass fit for the decay channel $B^0 \to D\pi, D \to K\pi\pi$ was already discussed in detail in section 5.2. The following sections will therefore focus on the fit of the B^0 lifetime in section 7.2 and proceed to the fit of Δm_d in section 7.3.

7.2 Lifetime Fit with a Neural Network Selection

7.2.1 Likelihood Function

The unbinned maximum likelihood fit used to extract the value of the B^0 lifetime τ is performed simultaneously in the mass and lifetime space using the negative log likelihood function

$$\mathcal{L} = -2 \sum_i \ln\Big[\big(1 - f_B\big) \cdot P_S(t_i, \sigma_{t,i}) \cdot P_S(m_i) + f_B \cdot P_B(t_i) \cdot P_B(m_i)\Big] \tag{7.1}$$

Here $f_B = \frac{N_{BG}}{N_{BG}+N_{Sig}}$ denotes the relative fraction of background events in the fit range used to weight the contributions in lifetime and mass space of signal events $P_S(t_i, \sigma_{t,i}) \cdot P_S(m_i)$ and background events $P_B(t_i) \cdot P_B(m_i)$.
The parameter f_B is shared by mass and lifetime spaces, but can be constrained by the fit in mass space only. The invariant mass distributions of signal and background contributions can be separated well, while in lifetime space the distributions of signal and background events have a very similar shape. The index i iterates over all candidates in the data samples serving as input to the fit.
The analytic expressions used for $P_{S,B}(m_i)$ and $P_S(t_i, \sigma_{t,i})$ have been introduced in sections 5.2.1 and 5.3.2. $P_B(t_i)$ is a parametric description of the shape of the lifetime distribution of combinatorial background events. It consists of a sum of two exponential functions convolved with Gaussian functions:

$$P_B(t) = N_B \left(f_1\, e^{-\frac{t}{\tau_1}} \otimes \mathcal{G}(t - t', \mu_1, \sigma_1) \; + \; (1 - f_1)\, e^{-\frac{t}{\tau_2}} \otimes \mathcal{G}(t - t', \mu_2, \sigma_2) \right) \tag{7.2}$$

where f_1, τ_1, τ_2, σ_1, σ_1/σ_2, $\mu 1$ and $\mu_1 - \mu_2$ are used as fit parameters. N_B normalizes the function to one. For the combined fit in mass and lifetime space, the mass range

5.2 GeV to 5.8 GeV is used. This reduces the contribution of partially reconstructed B^0 decays in the lower sideband region to a minimum so that their contributions can be neglected in the lifetime fit.
The probability density function used for signal events in lifetime space corresponds to eq. 5.21, using an an uncertainty $\sigma_{t,i}$ on an event–by–event basis. The initial values of the parameters of $P_B(t)$ are obtained from a template fit, see appendix E. During the unbinned maximum likelihood fit these parameters are treated as free parameters, only σ_1/σ_2 is fixed to the value obtained from the template fit. The fit converges to values for the background contribution that are very similar to those obtained from the template fit. Templates for the lifetime distributions of the contributions from $\Lambda_b \to \Lambda_c\pi$ and $B_s \to D_s\pi$ are added to $P_S(t_i, \sigma_{t,i})$ with the same relative fractions $f_{\Lambda_b \text{ as } B^0}$ and $f_{B_s \text{ as } B^0}$ that are used in the fit of the invariant mass spectrum (see section 5.2.1). The shapes of these templates can be found as well in appendix E.
For the combined fit in mass and lifetime space the mean of the two Gaussian functions and the width of the narrower of the two Gaussian functions modeling the signal peak are free parameters in mass space. All other parameters in mass space are fitted first by a binned likelihood fit in the wide invariant mass range from 4.7 to 6.0 GeV and a subsequent binned likelihood fit in the narrow invariant mass range from 5.2 to 5.8 GeV and then kept fixed during the unbinned maximum likelihood fit. The parameters for the efficiency function $\epsilon(t)$ are also fixed once they have been determined for a specific selection cut (see following section). In lifetime space, the B meson lifetime τ is a free parameter, as well as the background fraction f_B, shared by both spaces.

7.2.2 Determining the Efficiency Curve

The shape of the efficiency function $\epsilon(t)$ is determined according to eq. 5.24 using the proper decay time distribution obtained from a simulation of B meson decays, taking into account the bias introduced at trigger level and by selection cuts. This distribution is obtained after applying the same kind of cut or neural network selection on the sample of simulated events as is applied to the data sample that serves as input for the unbinned maximum likelihood fit. To make sure that the simulated events describe the signal distribution in data also after applying a cut on the neural network output, a comparison of the distributions of all quantities involved in the network training has to be performed[1]. If the distributions differ, one cannot expect that the efficiency curve derived after applying all preselection cuts to simulated events corresponds to the efficiency curve observed in real data when applying the same kind of preselection. Figure 7.2 compares the neural network output obtained for the

[1]For the cut based selection such a comparison was already performed (see appendix D) to decide which variables are well described by the simulation and are suited to serve as training variables in the network training.

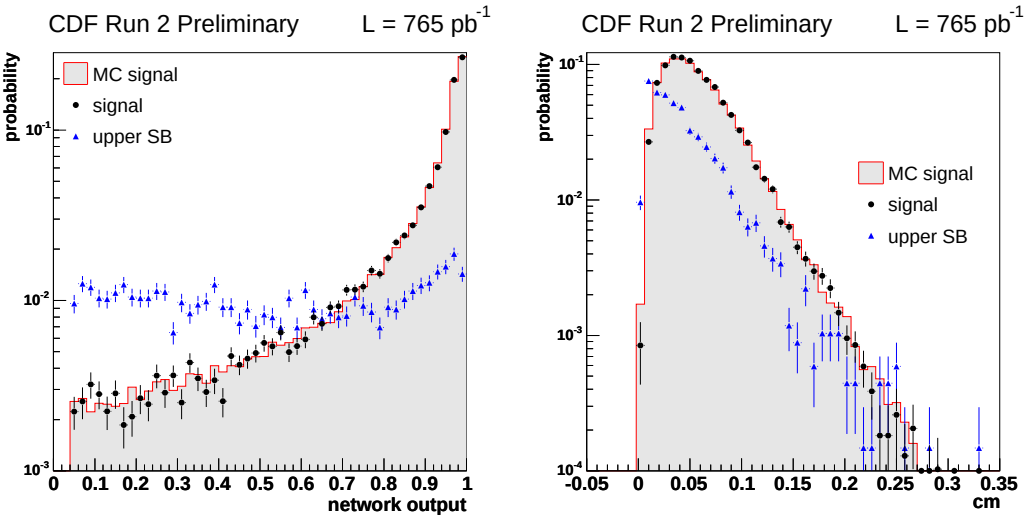


Figure 7.2: Comparison of the distributions of the neural network output (left) and proper decay time (right) for sideband subtracted signal events in data and simulated events (MC) after applying the cut on the network output that maximizes $\frac{S}{\sqrt{S+B}}$. As a reference, also the distributions from the upper sideband in data are displayed.

sample of simulated events and the proper decay time distribution of signal events in data after performing a sideband subtraction. In both cases the distributions obtained from simulated events and data agree very well. Comparisons of all variables used in the network training and details on the sideband subtraction procedure can be found in appendix D.4.

To determine the lifetime efficiency curve $\epsilon(t)$ after a neural network selection, the sample of simulated events has to be classified and the network cut that is applied to the data sample has to be applied to the sample of simulated events as well. The proper decay time distribution of selected simulated events $g(t)$ is filled into a histogram and divided by the integral of the function describing the unbiased lifetime distribution over the width of each bin of the proper decay time histogram according to eq. 5.24 :

$$\epsilon(t) = \frac{g(t) \text{ after trigger and cuts}}{\sum_i \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes \mathcal{G}(t - t', \sigma_{t,i})}$$

The resulting distribution is fitted using a phenomenological parametrization of $\epsilon(t)$ [53]

$$\epsilon(t) = \sum_{j=0}^{2} a_j \, (t - \xi_j)^2 \, e^{-t/\tau_j} \, \theta(t - \xi_j) \tag{7.3}$$

where $\theta(t - \xi_j)$ represents the Heavyside step function $\theta(x) = 1$ if $x > 0$ and $\theta(x) = 0$

else. This parametric form of $\epsilon(t)$ has the advantage that the normalization factor $N_S(t_i, \xi_i, \sigma_{t,i})$ can be calculated analytically (see appendix C). $N_S(t_i, \xi_i, \sigma_{t,i})$ keeps the probability density function used for the lifetime and mixing fits (eqs. 5.21 and 5.25) normalized to one for each event during the unbinned maximum likelihood fit,. Figure 7.3 shows the resulting efficiency curves $\epsilon(t)$ after fitting the parametrization from eq. 7.3 to the distributions obtained from data using a cut based selection and the neural network selection. The two efficiency curves look very similar.

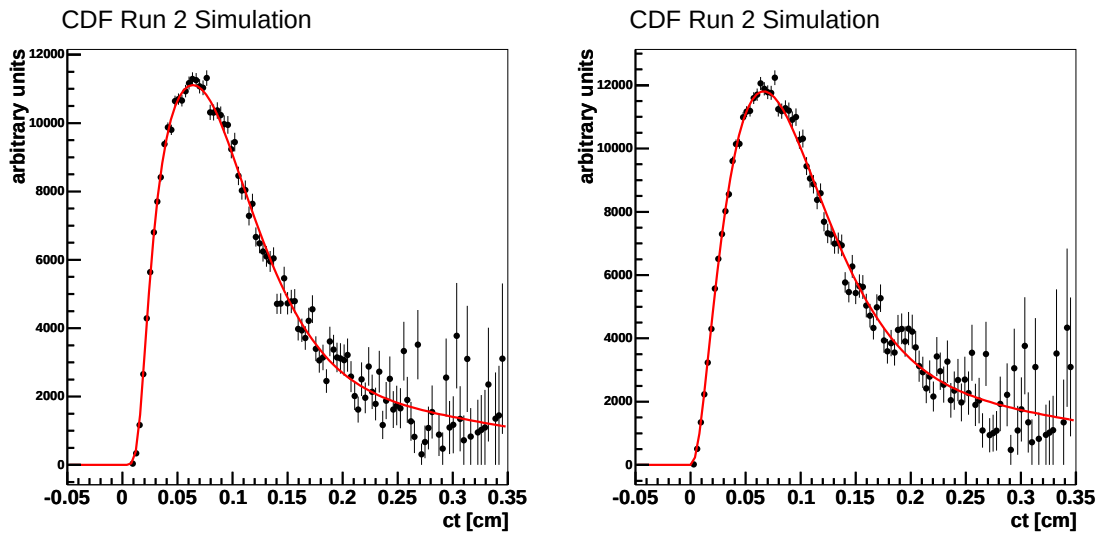


Figure 7.3: Fit of the parametrized $\epsilon(t)$ to the distributions obtained after applying a cut based selection (left) and a selection based on a neural network classification (right).

7.2.3 Results of the Lifetime Fit

The parameters obtained from the efficiency fit are used as input for the unbinned maximum likelihood fit, which is performed in parallel in mass and lifetime space. The negative log likelihood expression given in eq. 7.1 is minimized using the `MINUIT` [47] package. Figure 7.4 shows the projections of the fit function in lifetime space using the fit results obtained from the unbinned maximum likelihood fit for the cut based and the network based selections. Figure 7.5 shows the corresponding projections of the fit function in mass space.

Table 7.1 summarizes the results obtained using the cut based selection and a neural network selection. f_B is the common background fraction found in mass and lifetime space and the parameter *"signal width"* denotes the width of the narrower of the

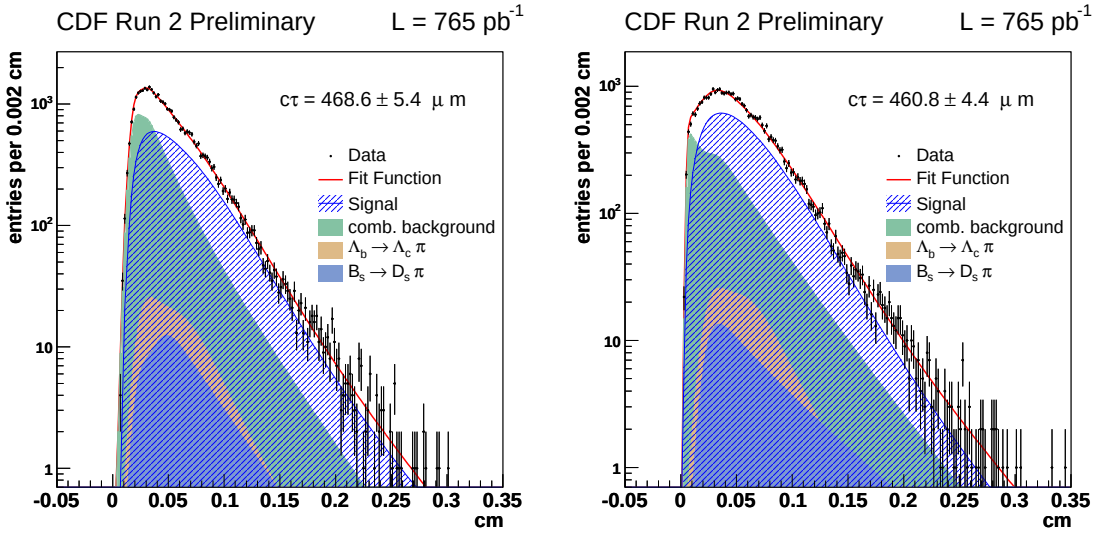


Figure 7.4: Projections of the fit function in lifetime space using the results obtained from the unbinned maximum likelihood fit using data selected by sequential cuts (left) and a selection based on a neural network classification (right).

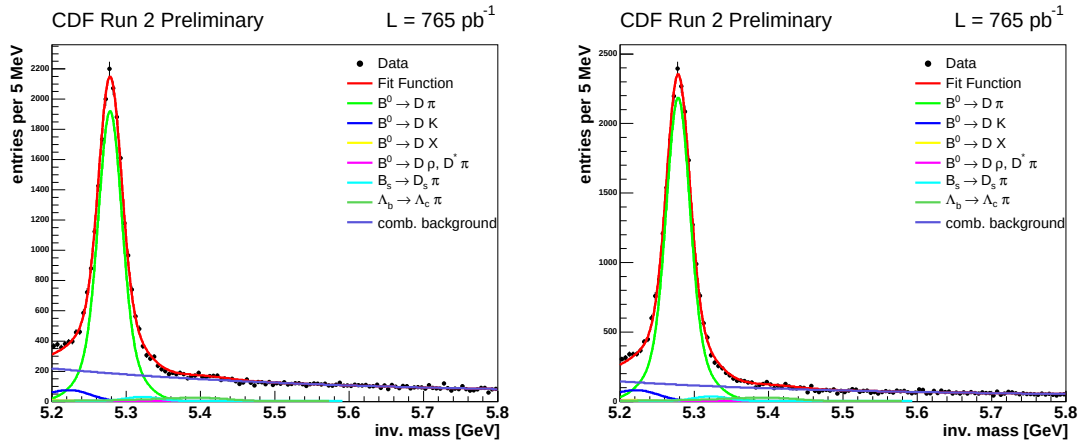


Figure 7.5: Projections of the fit function in mass space using the results obtained from the unbinned maximum likelihood fit using data selected by sequential cuts (left) and a selection based on a neural network classification (right).

Parameter	Result (cuts)	Result (network)
f_B	0.43172 ± 0.00368	0.29643 ± 0.00356
signal ct [μm]	468.62 ± 5.39	460.77 ± 4.45
signal mass [GeV]	5.2784 ± 0.0002	5.2783 ± 0.0002
signal width [GeV]	0.01544 ± 0.00015	0.01554 ± 0.00014
bg f_1	0.32271 ± 0.04647	0.33924 ± 0.04606
bg $c\tau_1$ [μm]	0.03248 ± 0.00127	0.04045 ± 0.00123
bg μ_1 [GeV]	0.02818 ± 0.00101	0.02605 ± 0.00148
bg σ_1 [GeV]	0.00361 ± 0.00014	0.00743 ± 0.00053
bg $c\tau_2$ [μm]	0.01671 ± 0.00133	0.02695 ± 0.00225
bg $\mu_1 - \mu_2$ [GeV]	-0.01150 ± 0.00084	-0.02051 ± 0.00144

Table 7.1: Summary of fit results obtained from the unbinned maximum likelihood fit of the mass and lifetime distributions after a cut based selection and after a neural network selection.

two Gaussian functions used as signal parametrization. All parameter names starting with *"bg"* denote parameters that describe the combinatorial background's proper decay time distribution (cf. eq. 7.2). The result obtained for the B^0 lifetime after the cut based selection is $c\tau = 468.2 \pm 5.4\,\mu m$, the neural network selection gives $c\tau = 460.8 \pm 4.5\,\mu m$. Both values are in good agreement with the world average value of $c\tau(B^0) = 459.3 \pm 2.0\,\mu m$ [15]. This result demonstrates that the fit model, incorporating the finite experimental resolution and the bias in the lifetime distribution due to trigger and selection cuts, is capable of extracting the correct lifetime from the underlying data sample. The bias in the lifetime distribution that is introduced by a neural network selection is similar to the bias observed after a cut based selection and can be incorporated by deriving the efficiency $\epsilon(t)$ using simulated events.

7.3 Fit of Δm_d

The negative log likelihood function used for the unbinned maximum likelihood fit of Δm_d is very similar to the expression used for fit of the B^0 lifetime in eq. 7.1. In addition to the description of signal and background distributions in mass and lifetime space, a term describing the $B^0 - \bar{B}^0$ mixing enters the probability density function that is used for the fit of Δm_d. The negative log likelihood function used for the

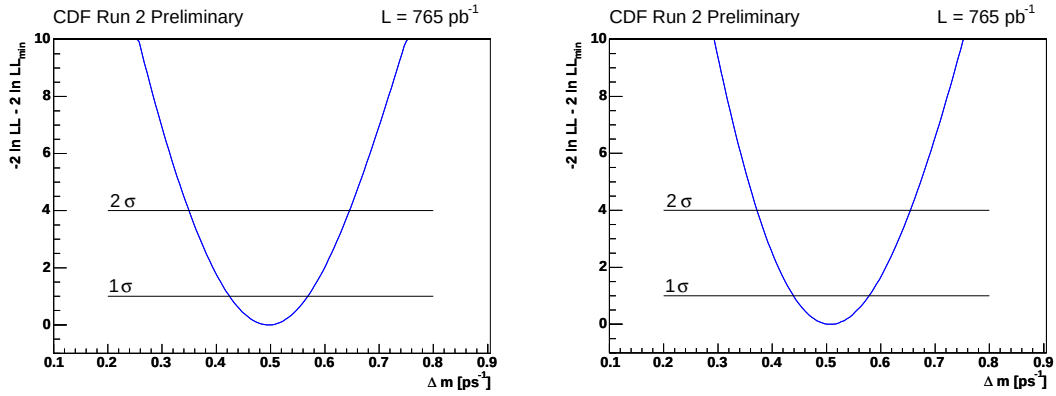


Figure 7.6: Likelihood function as a function of Δm_d, evaluated around the minimum obtained when using a cut based selection (left) and a selection based on a neural network classification (right) to provide the events that are used in the unbinned maximum likelihood fit.

minimization can be written as

$$\mathcal{L} = -2 \sum_i \ln\Big[\big(1 - f_B\big) \cdot P_S(t_i, \xi_i, \sigma_{t,i}) \cdot P_S(m_i) + f_B \cdot P_B(t_i) \cdot P_B(m_i)\Big] \tag{7.4}$$

where $P_S(t_i, \xi_i, \sigma_{t,i})$ denotes the lifetime distribution including the mixing term (see eq. 5.26), which contains the tagging decision ξ as an additional measured quantity. For this study only tagging decisions provided by the jet charge tagger [56] are used. The full analytical expression for $P_S(t_i, \xi_i, \sigma_{t,i})$ can be found in appendix C. During the minimization all parameters in lifetime and mass space are fixed to the values obtained after the unbinned maximum likelihood fit of the B^0 lifetime (see table 7.1). Only Δm_d remains as a free parameter. The results for Δm_d obtained after performing a cut based or a neural network selection are

$$\Delta m_d = 0.496 \pm 0.072 \text{ ps}^{-1} \qquad \text{after a cut based selection}$$
$$\Delta m_d = 0.507 \pm 0.069 \text{ ps}^{-1} \qquad \text{after a network based selection.}$$

The values obtained for Δm_d are in good agreement with the world average value of $\Delta m_d = 0.507 \pm 0.004$ ps^{-1} [10]. Figure 7.6 shows the negative log likelihood function for Δm_d evaluated in the region from 0.1 ps^{-1} to 0.9 ps^{-1}. Statistical uncertainties are quoted only. No systematic studies are performed as this is not intended to be a measurement of Δm_d but rather a check, demonstrating the potential to use the neural network selection for a mixing fit and the functionality of the fit framework.

7.4 Amplitude Scan and Fit of Δm_s

The previous section used data samples containing the decay $B^0 \to D\pi, D \to K\pi\pi$ obtained after applying a cut based optimization procedure and a neural network classification and tagging decisions provided by the jet charge tagger to perform a fit of Δm_d. The data sample obtained for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ is considerably smaller and extracting the oscillation frequency Δm_s is more difficult, as Δm_s is significantly larger than Δm_d. The measurement of Δm_s performed by CDF II [11] uses a set of five B_s decay modes and combines the information of four flavor taggers, so a check of the existing measurement or even an improvement, using a neural network selection, requires the use of more decay modes and the combination of several tagging algorithms.
Nevertheless, the decay mode $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ is the most important single mode for a Δm_s measurement, as this channel provides the largest number of signal events ($\simeq 40\%$ of the sum of the number of signal events of all five decay modes) and the best signal over background ratio of all decay modes. Therefore in this section a comparison of the amplitude scan method (see section 5.5.1) and an unbinned maximum likelihood fit of Δm_s performed on the data sample of the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ obtained after a cut based optimization procedure and a neural network classification is presented. This comparison is intended to demonstrate that the use of the neural network selection is capable of improving the existing Δm_s measurement, using the cut based optimization procedure.
The amplitude scan is performed in close analogy to the fit of Δm_d presented in the previous section. First all parameters in mass and lifetime space are fitted using the sequence of binned and unbinned likelihood fits introduced in section 7.1. All parameters in mass and lifetime space, including the B_s lifetime and the background fraction, are fixed to the values obtained from the fit sequence. The amplitude scan consists of a series of unbinned maximum likelihood fits where in each fit Δm_s is fixed to a certain value and the amplitude A (see eq. 5.28) is a free parameter. Figure 7.7 compares the amplitude scan obtained for a range of Δm_s from 1 to 21 ps^{-1} for the cut based selection and the neural network based selection. Both plots looks similar, as the samples obtained after both selections show a significant overlap (see figure 6.10). The sensitivity for Δm_s is obtained from the intersection of the dotted line, denoting values of $1.65\,\sigma_A$ derived from the blind amplitude scan (see section 5.5.1), with the line $A = 1$. Compared to the cut based selection, the sensitivity for Δm_s is improved from $\simeq 17$ ps^{-1} to $\simeq 18.5$ ps^{-1} using the neural network based selection. Also the 95 % confidence lower limit on Δm_s, indicated by a circle at the intersection of the $A + 1.65\,\sigma_A$ with the line $A = 1$ in both plots, improves from $\Delta m_s > 1.2$ ps^{-1} to $\Delta m_s > 9.8$ ps^{-1} using the neural network selection. The uncertainties quoted for the amplitude A in fig. 7.7 are statistical only and are derived from the unbinned maximum likelihood

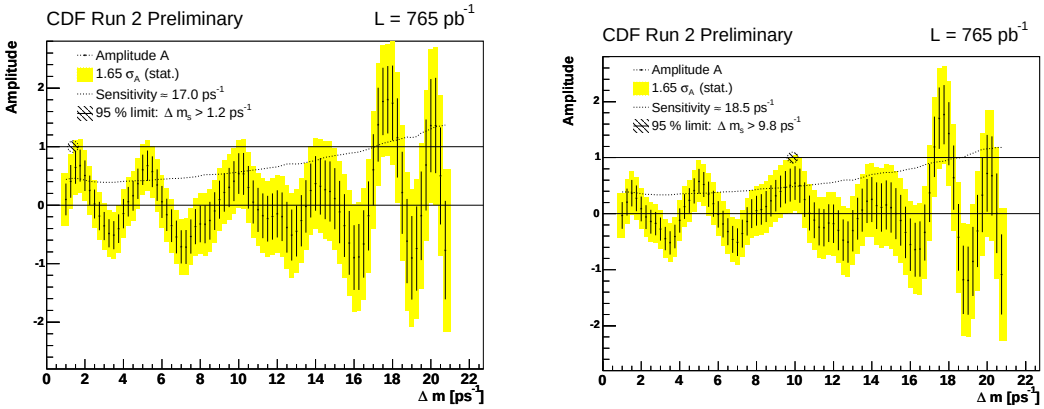


Figure 7.7: Amplitude scan for Δm_s obtained using a cut based selection (left) and a selection based on a neural network classification (right).

fit of A. The two amplitude scans are intended to compare the cut based and neural network based selection and demonstrate the relative improvement of the sensitivity achieved by a better candidate selection only. No treatment of systematic errors, most importantly the possibility of an incorrect dilution predicted by the tagging algorithm and the relative contribution of different background components, is performed, as they are not expected to behave significantly different for both types of selection. Another way to compare the data samples obtained using a neural network or cut based candidate selection is to use the samples as input for an unbinned maximum likelihood fit of Δm_s. The likelihood expression used in the fit, as well as the fit procedure, are identical to the fit of Δm_d performed in section 7.3. Figure 7.8 shows the likelihood function evaluated around the minimum obtained for Δm_s from an unbinned maximum likelihood fit. In both cases, the likelihood projection shows a complicated structure consisting of several local minima. This indicates that for a precise and significant measurement of Δm_s more decay channels and also more tagging information has to be included in the fit. However, in the range from 1 to 22 ps^{-1} both fits have their global minimum at $\simeq$ 17 ps^{-1}, where in the case of the sample selected by the neural network this minimum is still deeper compared to the local minima observed at $\simeq$ 14 ps^{-1} and $\simeq$ 20 ps^{-1}. The values for Δm_s obtained from both fits are:

$$\Delta m_s = 17.71 \pm 0.39\ ps^{-1} \qquad \text{after a cut based selection}$$
$$\Delta m_s = 17.64 \pm 0.22\ ps^{-1} \qquad \text{after a network based selection.}$$

Statistical uncertainties are quoted only. These numbers can't be interpreted as a final result for Δm_s, as no treatment of systematic uncertainties is performed.

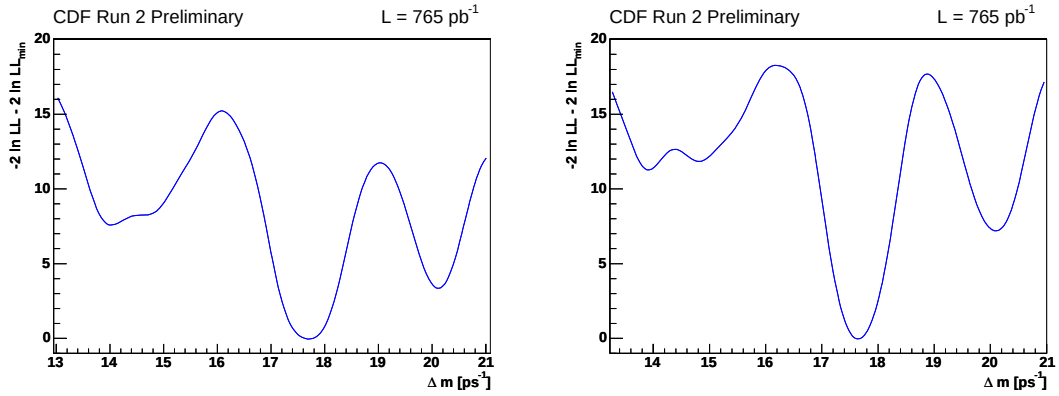


Figure 7.8: Likelihood function as a function of Δm_s, evaluated around the minimum obtained when using a cut based selection (left) and a selection based on a neural network classification (right) to provide the events that are used in the unbinned maximum likelihood fit.

The results of the comparison of the data sample selected by cuts to a data sample selected by a neural network are promising. In the amplitude scan as well as in the unbinned maximum likelihood fit the network selection could demonstrate that it will be capable of improving the existing Δm_s measurement when all decay channels and all the tagging information that is available are taken into account.

Chapter 8

Conclusion and Outlook

Using sophisticated neural network techniques, the event selection procedure for the decays of B mesons into exclusive final states has been improved significantly compared to the sequential cut based optimization procedure applied at CDF II up to now. The variables used in the selection procedure are combined optimally into a single discriminating variable, exploiting the additional information contained in the correlation between these variables. The networks are trained using simulated signal events and combinatorial background events from the upper sideband in data as training patterns. The novel technique developed in the course of this work improves the ongoing analysis of B_s mixing at CDF II and will substantially contribute to establishing a statistically significant values for Δm_s at the $5\,\sigma$ level.
Applying the network selection to the decay $B^0 \to D\pi, D \to K\pi\pi$ yields a signal significance of 131.5 $\pm$ 0.9 , which corresponds to an improvement of 11 % compared to the cut based optimization procedure that was applied at CDF II up to now. This improvement is equivalent to adding 24 % more data to the analysis. Extending the method to the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ gives $\frac{S}{\sqrt{S+B}} = 35.5 \pm 0.7$, an improvement of 16 % compared to the cut based optimization procedure, which corresponds to adding 35 % more data to the analysis. Taking into account the ct–resolution of the selected events, this leads to an expected overall improvement of 14 % of the significance of the B_s mixing measurement in this decay channel. The dataset that is used for both decay channels corresponds to an integrated luminosity of 765 ps^{-1}.
Comparing the data samples in the decay chain $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ selected by cuts or a neural network in an amplitude scan and an unbinned maximum likelihood fit of Δm_s confirms that the neural network selection leads to a significant improvement of the existing measurement of Δm_s in the near future.
In addition, a fit framework allowing the measurement of the lifetime of exclusively reconstructed B mesons from a data sample with a lifetime distribution that is biased

due to trigger and selection cuts has been implemented. The framework allows to determine the mixing frequency Δm_d (Δm_s) of B^0 (B_s) mesons by incorporating tag decisions provided by flavor tagging algorithms. Beyond the application in a measurement of Δm_s, in future the fit framework can be used for the calibration of advanced tagging algorithms for measurements of the CP–asymmetry in the B_s system.
As a check of both improved selection procedure and fitter framework, the lifetime and mixing frequency of the B^0 meson in the final state $B^0 \to D\pi, D \to K\pi\pi$ was determined by performing either a cut based selection or a neural network selection:

$$c\tau(B^0) = 468.2 \pm 5.4\,\mu m$$
$$\Delta m_d = 0.496 \pm 0.072\ ps^{-1} \qquad \text{after a cut based selection}$$

$$c\tau(B^0) = 460.8 \pm 4.5\,\mu m$$
$$\Delta m_d = 0.507 \pm 0.069\ ps^{-1} \qquad \text{after a network based selection.}$$

Both results are in good agreement with the world average values of $c\tau(B^0) = 459.3 \pm 2.0\,\mu m$ [15] and $\Delta m_d = 0.507 \pm 0.004$ ps^{-1} [10].
Also an unbinned maximum likelihood fit of Δm_s was performed using the decay channel $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$, resulting in

$$\Delta m_s = 17.71 \pm 0.39 \text{ ps}^{-1} \qquad \text{after a cut based selection}$$
$$\Delta m_s = 17.64 \pm 0.22 \text{ ps}^{-1} \qquad \text{after a network based selection.}$$

Statistical uncertainties are quoted only. These numbers can not be interpreted as a final result for Δm_s, as no treatment of systematic uncertainties is performed.
The improved signal selection using neural networks can be easily transferred to other exclusive final states and enhance any type of analyses involving exclusively reconstructed B Meson decays. For an application of the neural network selection in the context of a Δm_s measurement, all final states contributing to the measurement have to be individually optimized with neural networks. First results [66] show that a significantly larger improvement of the signal significance $\frac{S}{\sqrt{S+B}}$ can still be achieved for other B_s decay channels like $B_s \to D_s\pi\pi\pi$. These decays are kinematically more complex, involve subresonances and contain, due to six instead of four tracks in the final state, a larger fraction of combinatorial background. Extending the network selection to all decay modes relevant for the B_s mixing measurement will therefore provide an improvement of the existing Δm_s measurement in the near future and contribute significantly to the discovery of B_s oscillations at CDF II.

Appendix A

Definition of Variables

The following variable definitions are used throughout the work presented in this thesis. The pion sub– and superscripts are used to distinguish the pion π_B originating directly from the B meson decay from the two pions π_D^1 and π_D^2 from the subsequent D meson decay $D \to K\pi\pi$.

$|d_0(B)|$ Impact parameter of the B meson

$L_{xy}(B)$ Decay length of the B meson defined as $L_{xy} = \frac{\vec{r}\cdot\vec{p}}{p_t}$, where $\vec{r}$ is the displacement of the decay position in the (x–y) plane, $\vec{p}$ the momentum vector of the particle in (x–y) and $p_t = |\vec{p}|$

$L_{xy}/\sigma_{L_{xy}}(B)$ Decay length significance of the B meson

$\chi^2_{3D}(B)$ Three–dimensional χ^2 of the B vertex fit

$\chi^2_{r\phi}(B)$ Two–dimensional χ^2 of the B vertex fit

$p_t(B)$ Transverse momentum of the B meson

$|d_0(D)|$ Impact parameter of the D meson

$\chi^2_{3D}(D)$ Three–dimensional χ^2 of the D vertex fit

$\chi^2_{r\phi}(D)$ Two–dimensional χ^2 of the D vertex fit

$L_{xy}/\sigma_{L_{xy}}(D)$ Decay length significance of the D meson

$L_{xy}(D)$ Decay length of the D meson

$p_t(D)$ Transverse momentum of the D meson

$L_{xy}(B \to D)$ Decay length of the D meson with respect to the B meson

$\min(m_{K,\pi_D^1}, m_{K,\pi_D^2})$ Minimum of the two invariant masses calculated from the four momenta of either pion from the decay $D \to K\pi\pi$ and the four momentum of the kaon

$\max(m_{K,\pi_D^1}, m_{K,\pi_D^2})$ Maximum of the two invariant masses calculated from the four momenta of either pion from the decay $D \to K\pi\pi$ and the four momentum of the kaon

$\sphericalangle(\vec{p}(B), \vec{p}(\pi_B))_{\mathrm{cms}B}$ Cosine of the angle between the three–momentum of the B meson $\vec{p}(B)$ and the three–momentum $\vec{p}(\pi_B)$ of the π_B boosted to the B mesons rest frame

$\Delta R(D, \pi_B)$ Opening angle (in sterad) between the D meson and the π from the decay $B^0 \to D\pi$

$p_t(\pi_B)$ Transverse momentum of the π_B

$p_t(K)$ Transverse momentum of the kaon from the decay $D \to K\pi\pi$

$p_t(\pi_D^1)$ Transverse momentum of the first pion from the decay $D \to K\pi\pi$

$p_t(\pi_D^2)$ Transverse momentum of the second pion from the decay $D \to K\pi\pi$

$|d_0(\pi_B)|$ Impact parameter of the π_B

$|d_0(K)|$ Impact parameter of the kaon from the decay $D \to K\pi\pi$

$|d_0(\pi_D^1)|$ Impact parameter of the first pion from the decay $D \to K\pi\pi$

$|d_0(\pi_D^2)|$ Impact parameter of the second pion from the decay $D \to K\pi\pi$

lifetime–signed $d_0(\pi_B)$ Lifetime signed impact parameter of the π_B (see fig. A.1)

lifetime–signed $d_0(K)$ Lifetime signed impact parameter of the kaon from the decay $D \to K\pi\pi$

lifetime–signed $d_0(\pi_D^1)$ Lifetime signed impact parameter of the first pion from the decay $D \to K\pi\pi$

lifetime–signed $d_0(\pi_D^2)$ Lifetime signed impact parameter of the second pion from the decay $D \to K\pi\pi$

min $|d_0|$ Minimum impact parameter of all four tracks

min $\frac{|d_0|}{\sigma_{d0}}$ Minimum impact parameter significance of all four tracks

min p_t Minimum transverse momentum of all four tracks

helicity angle K^1 Cosine of the angle between the three–momentum of the B_s meson $\vec{p}(B)$ and the three–momentum $\vec{p}(K^1)$ of the first of the two kaons from the decay $\phi \to KK$ boosted to the B_s mesons rest frame (the "helicity angle")

$m_{\phi \to KK}$ Invariant mass of the decay $\phi \to KK$

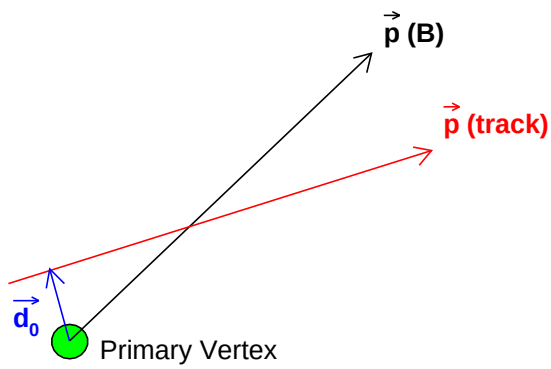


Figure A.1: Definition of the lifetime–signed impact parameter for a track in an event with a B decay. $d_0^{lts} = |\vec{d_0}|\ \mathrm{signum}(\vec{d_0}\,\vec{p_B})$

Appendix B

Details of the Variable Significance Table

Tables 6.1, 6.2 and 6.3 show examples for variable significance tables calculated as part of the NeuroBayes® preprocessing [34][67][68]. Taking into account all $N-1$ training variables and the training target (i.e. signal or background) the $N \times N$ covariance matrix C is calculated after the variables' individual preprocessing. The entries C_{ij} are calculated for all pairs of variables (including the target) x_i, x_j

$$C_{ij} = \frac{1}{n} \sum_{\text{events}} \left(x_i - < x_i >\right) \cdot \left(x_j - < x_j >\right) \tag{B.1}$$

where $< x >$ denotes the average or expectation value of x and n is the number of training events. The correlation between two variables can be expressed by the entries of the correlation matrix $\rho_{ij} = \frac{C_{ij}}{\sqrt{V[x_i]}\sqrt{V[x_j]}}$ where $V[x_i]$ denotes the variables' variance and ρ_{ij} the correlation coefficient.

The NeuroBayes® preprocessing then decorrelates all training variables by appropriate rotations of the $(N-1) \times (N-1)$ correlation matrix of the training variables. The rotations performed to diagonalize the training variables' correlation matrix are performed simultaneously on the vector $\vec{\rho} = (\rho_1, ..., \rho_{N-1})$ of each training variable's correlation to the training target[1]. The vector $\vec{\tilde{\rho}}$ obtained after the $(N-1) \times (N-1)$ correlation matrix of the training variables is diagonalized contains the correlations of the $N-1$ decorrelated training variables to the training target. The *total correlation* ρ_T of all training variables to the training target then is ${\rho_T}^2 = \sum_{i=1}^{N-1} \tilde{\rho}_i^{\,2}$.
To calculate the variable significance table (table B.1 shows an example for such a variable table obtained after a training with four variables) for each of the $N-1$

[1]The vector $\vec{\rho}$ corresponds to the first row of the full $N \times N$ correlation matrix of all training variables and the target.

Rank	Name	Significance [σ]
1	$\lvert d_0(B)\rvert$	163.21
2	$L_{xy}/\sigma_{L_{xy}}(B)$	76.62
3	$\chi^2_{3D}(B)$	48.73
4	$p_t(\pi_B)$	41.43

Table B.1: Example for a variable significance table obtained after the NeuroBayes® preprocessing.

variables x_i the *loss of total correlation*

$$\rho^2_{x_i} = \rho^2_{T,x_1,\ldots,x_{N-1}} - \rho^2_{T,x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_{N-1}} \tag{B.2}$$

is calculated if this variable was removed from the set of training variables. The variable x_i where the smallest loss of total correlation occurs is then removed from the set of variables. In table B.1 this is variable $p_t(\pi_B)$, at the bottom of the table. The significance S of this variable, which is quoted in the right column of table B.1, is then obtained from $S = \rho_{x_i} \cdot \sqrt{n}$ where n is the number of events in the training sample. The significance S of a variable serves as a measure of the additional total correlation to the target that is introduced by adding this variable to the remaining set of $N-2$ training variables. This procedure is repeated with the set of $N-2$ variables until the significance of all variables is calculated. As the significance of the k–th variable in the table depends only on the $k-1$ remaining variables, the variable table is not necessarily ordered monotonicly in significance.

Appendix C

Derivation of analytic expressions used in the mixing fit

The probability density function describing a mixing signal that is sculpted by an efficiency curve $\epsilon(t)$ is given in section 5.5 by equation 5.25

$$P_S(t_i, \xi_i, \sigma_{t,i}) = \frac{1}{N_S(t, \xi, \sigma_t)} \cdot \Big(\frac{1 + \xi A D \cos(\Delta m\, t)}{1 + |\xi|} \frac{1}{\tau} e^{-\frac{t}{\tau}} \Big) \otimes \mathcal{G}(t - t', \sigma_t) \cdot \epsilon(t) \quad \text{(C.1)}$$

A denotes the amplitude factor encountered in section 5.5.1. Here again $\otimes$ is used as abbreviation for the convolution integral

$$\begin{aligned} f(t_i, \xi_i, \sigma_{t,i}) &= \frac{1}{N_S(t, \xi, \sigma_t)} \cdot \Big(\frac{1 + \xi D \cos(\Delta m\, t)}{1 + |\xi|} \frac{1}{\tau} e^{-\frac{t}{\tau}} \Big) \otimes \mathcal{G}(t - t', \sigma_t) \\ &= \frac{1}{N_S(t, \xi, \sigma_t)} \cdot \int_{-\infty}^{\infty} \Big(\frac{1 + \xi D \cos(\Delta m\, t)}{1 + |\xi|} \frac{1}{\tau} e^{-\frac{t}{\tau}} \Big) \mathcal{G}(t - t', \sigma_t)\, dt' \end{aligned} \quad \text{(C.2)}$$

and $P_S(t_i, \xi_i, \sigma_{t,i}) = f(t_i, \xi_i, \sigma_{t,i}) \cdot \epsilon(t)$. It is advantageous to use $\cos(\Delta m\, t) = Re(e^{-i\Delta m t})$ and move to the complex plane for solving the integrals involved in the normalization of this function. The analytic solution of the integral defined by C.2 is given by

$$\begin{aligned} f(t_i, \xi_i, \sigma_{t,i}) = \frac{1}{2\tau} \frac{e^{-\frac{1}{\tau}(t - \frac{\sigma^2}{2\tau})}}{|\xi| + 1} \Bigg[&\mathrm{Erfc}\Big(\frac{\sigma^2 - t\tau}{\sqrt{2}\sigma\tau} \Big) \\ &+ \xi A D e^{-\frac{\sigma^2 \Delta m^2}{2}} \mathrm{Re}\Big\{ e^{-i\Delta m (t - \frac{\sigma^2}{\tau})} \mathrm{Erfc}\Big(\frac{\sigma^2 - t\tau}{\sqrt{2}\sigma\tau} + i \frac{\Delta m \sigma}{\sqrt{2}} \Big) \Big\} \Bigg] \end{aligned} \quad \text{(C.3)}$$

Using expression 7.3

$$\epsilon(t) = \sum_{j=0}^{2} a_j\, (t - \xi_j)^2\, e^{-t/\tau_j}\, \theta(t - \xi_j) \quad \text{(C.4)}$$

as parametrization of the lifetime efficiency curve ensures that the normalization factor $N_S(t_i, \xi_i, \sigma_{t,i})$ can be calculated analytically. Splitting C.3 into

$$f_1(t_i, \xi_i, \sigma_{t,i}) = \frac{1}{2\tau} \frac{e^{-\frac{1}{\tau}(t-\frac{\sigma^2}{2\tau})}}{|\xi|+1} \operatorname{Erfc}\Big(\frac{\sigma^2 - t\tau}{\sqrt{2}\sigma\tau}\Big) \tag{C.5}$$

and

$$f_2(t_i, \xi_i, \sigma_{t,i}) = \frac{1}{2\tau} \frac{e^{-\frac{1}{\tau}(t-\frac{\sigma^2}{2\tau})}}{|\xi|+1} \xi A D e^{-\frac{\sigma^2 \Delta m^2}{2}} \operatorname{Re}\Big\{e^{-i\Delta m(t-\frac{\sigma^2}{\tau})} \operatorname{Erfc}\Big(\frac{\sigma^2 - t\tau}{\sqrt{2}\sigma\tau} + i\frac{\Delta m \sigma}{\sqrt{2}}\Big)\Big\} \tag{C.6}$$

leads to

$$P_S(t_i, \xi_i, \sigma_{t,i}) = \frac{1}{N_S(t, \xi, \sigma_t)} \cdot \Big(f_1(t_i, \xi_i, \sigma_{t,i}) + f_2(t_i, \xi_i, \sigma_{t,i})\Big) \cdot \epsilon(t) \tag{C.7}$$

and the normalization constants ensuring $\int P_S(t_i, \xi_i, \sigma_{t,i}) dt = 1$ can be calculated from

$$N_S(t_i, \xi_i, \sigma_{t,i}) = \int_{-\infty}^{\infty} \Big(f_1(t_i, \xi_i, \sigma_{t,i}) + f_2(t_i, \xi_i, \sigma_{t,i})\Big) \cdot \epsilon(t)\, dt \tag{C.8}$$

The integral C.8 can be split into six parts for $\xi = 0$ or $\xi = \pm 1$ and sorted in powers of t^N $(N = 0, 1, 2)$, as $\epsilon(t)$ from C.4 can be written in the form

$$\epsilon(t) = \sum_{j=0}^{2} (d_j + c_j t + b_j t^2)\, e^{-a_j t}\, \theta(t + \frac{c_j}{2b_j}) \tag{C.9}$$

where $\alpha_j = 1/\tau_j$, $b_j = a_j$, $c_j = -2a_j\xi_j$ and $d_j = a_j\xi_j^2$. This leads to integrals for $N^j_{\xi N}(t_i, \xi_i, \sigma_{t,i})$ of the type

$$\begin{aligned}
\xi = 0,\, N = 0 :&\quad N^j_{00}(t_i, \xi_i, \sigma_{t,i}) = \int d_j\, e^{-\alpha_j t}\, f_1(t_i, \xi_i, \sigma_{t,i})\, dt \\
\xi = 0,\, N = 1 :&\quad N^j_{01}(t_i, \xi_i, \sigma_{t,i}) = \int c_j t\, e^{-\alpha_j t}\, f_1(t_i, \xi_i, \sigma_{t,i})\, dt \\
\xi = 0,\, N = 2 :&\quad N^j_{02}(t_i, \xi_i, \sigma_{t,i}) = \int b_j t^2\, e^{-\alpha_j t}\, f_1(t_i, \xi_i, \sigma_{t,i})\, dt \\
\xi = \pm 1,\, N = 0 :&\quad N^j_{10}(t_i, \xi_i, \sigma_{t,i}) = \int d_j\, e^{-\alpha_j t}\, f_2(t_i, \xi_i, \sigma_{t,i})\, dt \\
\xi = \pm 1,\, N = 1 :&\quad N^j_{11}(t_i, \xi_i, \sigma_{t,i}) = \int c_j t\, e^{-\alpha_j t}\, f_2(t_i, \xi_i, \sigma_{t,i})\, dt \\
\xi = \pm 1,\, N = 2 :&\quad N^j_{12}(t_i, \xi_i, \sigma_{t,i}) = \int b_j t^2\, e^{-\alpha_j t}\, f_2(t_i, \xi_i, \sigma_{t,i})\, dt
\end{aligned}$$

For each term $j = 0, 1, 2$ in $\epsilon(t)$ the six contributions $N^j_{\xi N}(t_i, \xi_i, \sigma_{t,i})$ add up to $N^j_S(t_i, \xi_i, \sigma_{t,i})$. The overall normalization of $P_S(t_i, \xi_i, \sigma_{t,i})$ is then given by

$$N_S(t_i, \xi_i, \sigma_{t,i}) = \sum_{j=0}^{2} N^j_S(t_i, \xi_i, \sigma_{t,i}) = \sum_{j=0}^{2}\sum_{N=0}^{2}\sum_{i=0}^{1} N^j_{iN}(t_i, \xi_i, \sigma_{t,i}) \tag{C.10}$$

where j sums over the three components of $\epsilon(t)$, the sum over N adds the different powers in t for each part of $\epsilon(t)$ and the sum over i puts together the normalizations for the lifetime and mixing terms. and show tests of the analytical implementation in C++ program code with a numerical integration algorithm[47].

C.1 Analytical Results of the Normalization Integrals

This section summarizes the analytical results of each integral $N^j_{\xi N}(t_i, \xi_i, \sigma_{t,i})$ obtained by the help of [69] The integration limits and the index j for the different terms of $\epsilon(t)$ are omitted.

$$N_{00} = \frac{d}{2(|\xi|+1)\tau(1+\alpha\tau)}\left(e^{-(\alpha+1/\tau)t}\tau\left[e^{\frac{\alpha^2\sigma^2}{2}+\alpha t+t/\tau}\,\mathrm{Erf}\Big(\frac{\alpha\sigma^2+t}{\sqrt{2}\sigma}\Big) - e^{\frac{\sigma^2}{2\tau^2}}\,\mathrm{Erfc}\Big(\frac{\sigma^2-\tau t}{\sqrt{2}\tau\sigma}\Big)\right]\right)$$

$$\begin{aligned} N_{01} = {} & \frac{c\tau}{(|\xi|+1)\tau\sqrt{2\pi}\sigma(1+\alpha\tau)}\Bigg(-e^{-\alpha t-\frac{t^2}{2\sigma^2}}\,\sigma^2 \;-\; \alpha e^{\frac{\alpha^2\sigma^2}{2}}\sqrt{\frac{2}{\pi}}\sigma^3\,\mathrm{Erf}\Big(\frac{\alpha\sigma^2+\tau}{\sqrt{2}\sigma}\Big) \\ & + \frac{1}{4}c\Bigg[\frac{\tau^2}{(1+\alpha\tau)^2}e^{\frac{\alpha^2\sigma^2}{2}}\,\mathrm{Erf}\Big(\frac{\alpha\sigma^2+\tau}{\sqrt{2}\sigma}\Big) \\ & - \frac{\tau(\tau+t+\alpha\tau t)}{(1+\alpha\tau)^2}\,e^{\frac{\sigma^2}{2\tau^2}-\frac{(1+\alpha\tau)t}{\tau}}\,\mathrm{Erfc}\Big(\frac{\sigma^2-\tau t}{\sqrt{2}\tau\sigma}\Big)\Bigg]\Bigg) \end{aligned}$$

$$
\begin{aligned}
N_{02} = {} & \frac{1}{2(|\xi|+1)\tau(1+\alpha\tau)^2}\Bigg(be^{-\alpha t-\frac{t^2}{2\sigma^2}}\sigma\tau\Bigg[-\sqrt{\frac{2}{\pi}}\Big(-\alpha\sigma^2+2\tau-\alpha^2\sigma^2\tau+t+\alpha\tau t\Big) \\
& + e^{\frac{(\alpha\sigma^2+t)^2}{2\sigma^2}}\sigma\Big(1+\alpha^2\sigma^2-\alpha\tau+\alpha^3\sigma^2\tau\Big)\,\mathrm{Erf}\Big(\frac{\alpha\sigma^2+\tau}{\sqrt{2}\sigma}\Big)\Bigg]\Bigg) \\
& + \frac{1}{4\tau(1+\alpha\tau)^3}\Bigg(b\tau\Bigg[2e^{\frac{\alpha^2\sigma^2}{2}}\tau^2\,\mathrm{Erf}\Big(\frac{\alpha\sigma^2+\tau}{\sqrt{2}\sigma}\Big)-e^{\frac{\sigma^2-2(1+\alpha\tau)t}{2\tau^2}} \\
& \cdot\Big(t^2+2\tau t(1+\alpha t)\Big)+\tau^2\Big(2+2\alpha t+\alpha^2t^2\Big)\,\mathrm{Erfc}\Big(\frac{\sigma^2-\tau t}{\sqrt{2}\tau\sigma}\Big)\Bigg]\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
N_{10} = {} & -d\,\xi AD\,\mathrm{Re}\Bigg\{-e^{-\frac{(1+2i\tau\Delta m+2\alpha\tau(1+i\tau\Delta m))\sigma^2}{2\tau^2}}\Bigg[e^{\frac{\Delta m^2\sigma^2}{2}+\frac{(1+\alpha\tau+i\tau\Delta m)((1+i\tau\Delta m)\sigma^2-\tau t)}{\tau^2}} \\
& \cdot\tau\Big(1+\alpha\tau+i\tau\Delta m\Big)\,\sigma\,\mathrm{Erfc}\Big(\frac{(1+i\tau\Delta m)\sigma^2-\tau t}{\sqrt{2}\tau\sigma}\Big) \\
& - e^{\frac{(1+\alpha\tau)(1+\alpha\tau+2i\tau\Delta m)\sigma^2}{\sigma\tau^2}}\sqrt{-\tau^2(1+\alpha\tau+i\tau\Delta m)^2\sigma^2} \\
& \cdot\,\mathrm{Erfi}\Big(\frac{\tau(1+\alpha\tau+i\tau\Delta m)(\alpha\sigma^2+t)}{\sqrt{2}\sqrt{-\tau^2(1+\alpha\tau+i\tau\Delta m)^2\sigma^2}}\Big)\Bigg]\Big/\Big(4\tau(1+\alpha\tau+i\tau\Delta m)^2\sigma\Big)\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
N_{11} = {} & c\,\xi AD\,\mathrm{Re}\Bigg\{-i\Big(-\sigma^2\,e^{-\alpha t-\frac{t^2}{2\sigma^2}}-\alpha\,e^{\frac{\alpha^2\sigma^2}{2}}\sqrt{\frac{\pi}{2}}\,\sigma^3\,\mathrm{Erf}\Big(\frac{\alpha\sigma^2+t}{\sqrt{2}\sigma}\Big)\Big) \\
& \Big/\Big(2\big(-i-i\alpha\tau+\tau\Delta m\big)\sqrt{2\pi}\sigma\Big) \\
& -\Bigg[-e^{-\frac{(-i+\tau\Delta m)^2\sigma^2+2\tau(1+\alpha\tau+i\tau\Delta m)}{2\tau^2}}\Bigg(\\
& -\tau\,e^{\frac{(1+\alpha\tau+i\tau\Delta m)((-1+\alpha\tau-i\tau\Delta m)\sigma^2+2\tau t)}{2\tau^2}}\,\mathrm{Erf}\Big(\frac{\alpha\sigma^2+t}{\sqrt{2}\sigma}\Big) \\
& +\big(\tau+t+\alpha\tau t+i\tau\Delta mt\big)\,\mathrm{Erfc}\Big(\frac{(1+i\tau\Delta m)\sigma^2-\tau t}{\sqrt{2}\tau\sigma}\Big)\Bigg)\Bigg] \\
& \Big/\Big(4(1+\alpha\tau+i\tau\Delta m)\Big)\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
N_{12} = b\,\xi AD\,\mathrm{Re}\Bigg\{\Bigg[&\sigma\, e^{-\alpha t-\frac{t^2}{2\sigma^2}}\Big(-\sqrt{2}\Big(-\alpha\sigma^2+t+\tau\big(2-\alpha^2\sigma^2+i\Delta mt+\alpha(-i\Delta m\sigma^2+t)\big)\Big)\\
&+e^{\frac{(\alpha\sigma^2+t)^2}{2\sigma^2}}\sqrt{\pi}\,\sigma\,\Big(1-\alpha\tau+i\tau\Delta m+\alpha^3\tau\sigma^2+\alpha^2(1+i\tau\Delta m)\sigma^2\Big)\\
&\mathrm{Erf}\Big(\frac{\alpha\sigma^2+t}{\sqrt{\pi}\sigma}\Big)\Big)\Bigg]\Big/\Big(4(1+\alpha\tau+i\tau\Delta m)^2\sqrt{2}\Big)\\
&+\Bigg[-e^{\frac{(-i+\tau\Delta m)^2\sigma^2+2\tau(1+\alpha\tau+i\tau\Delta m)t}{2\tau^2}}\,\tau\,\Big(t^2+2\tau t\big(1+\alpha t+i\Delta mt\big)\Big)\\
&+\tau^2\big(2+2i\Delta mt+\alpha^2t^2-\Delta m^2t^2+2\alpha t(1+i\Delta mt)\big)\\
&\mathrm{Erfc}\Big(\frac{(1+i\tau\Delta m)\sigma^2-\tau t}{\sqrt{2}\tau\sigma}\Big)\Bigg]\Big/\big(1+\alpha\tau+i\tau\Delta m\big)^3\\
&-2e^{\frac{\alpha^2\sigma^2}{2}}\,\tau^3\sigma\,\mathrm{Erfi}\left(\frac{\big(1+\alpha\tau+i\tau\Delta m\big)^3\big(\alpha\sigma^2+t\big)}{\sqrt{2}\sqrt{-\big(1+\alpha\tau+i\tau\Delta m\big)^6\sigma^2}}\right)\Bigg\}
\end{aligned}
$$

Here $\mathrm{Erfi}(z)$ denotes the complex error function defined as

$$\mathrm{Erfi}(z) = -i\,\mathrm{Erf}(\mathrm{iz}) \tag{C.11}$$

The evaluation of the error function $\mathrm{Erf}(z)$ with complex argument z is based on a asymptotic series expansion suggested in [70] defining

$$w(iz) = e^{z^2}\,\mathrm{Erf}(\mathrm{z}) \tag{C.12}$$

and using the expansion for w

$$w(iz) = \frac{1}{\sqrt{\pi}z}\left[1+\sum_{m=1}^{\infty}(-1)^m\frac{(2m-1)!!}{(2z^2)^m}\right] \tag{C.13}$$

C.2 Numerical Tests of the Analytic Implementation

The integrals introduced in section C.1 are rather complicated expressions and an implementation in C++ source code has to be checked for possible programming mistakes or other types of inconsistencies. This check is performed comparing the implementation of the analytical solutions of the integrals to a numerical integration

performed using the ROOT package[47]. Figure C.1 and C.2 show this comparison for the normalization of $f_1(t_i, \xi_i, \sigma_{t,i})$ and $f_2(t_i, \xi_i, \sigma_{t,i})$ respectively for typical values of all parameters. In each case the difference between the implementation of the analytical function and the numerical integral is $< 10^{-6}$, which can be attributed to the limited precision of the numerical integration routine.

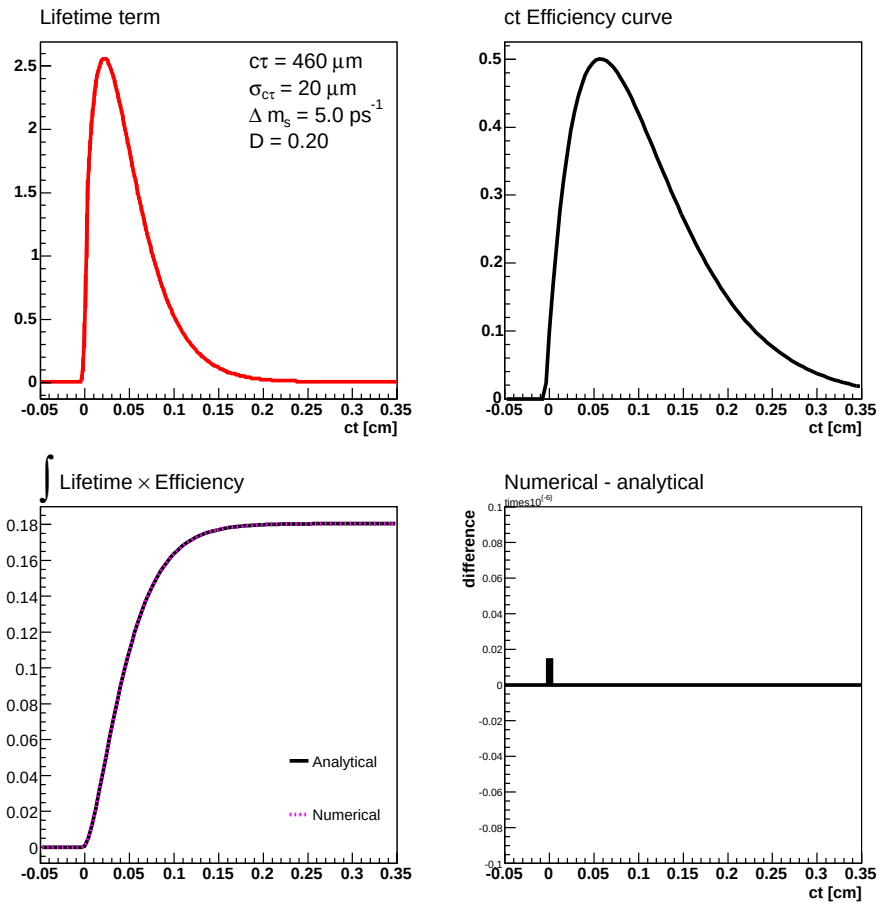


Figure C.1: Comparison of the implemented analytical solution of the sum of integrals $N_{00}+N_{01}+N_{02}$, this corresponds to the normalization of the function $f_1(t_i, \xi_i, \sigma_{t,i})$, to a numerical integration. The top left panel shows $f_1(t_i, \xi_i, \sigma_{t,i})$ for some typical values of $c\tau$ and σ_t. The top right panel shows the efficiency curve that is used, the bottom left panel shows the analytical and numerical integrals $\int_{-0.05}^{x} f_1(t_i, \xi_i, \sigma_{t,i})$ superimposed, where the x–axis denotes the upper limit used for the integration, and the bottom right panel the difference of the two.

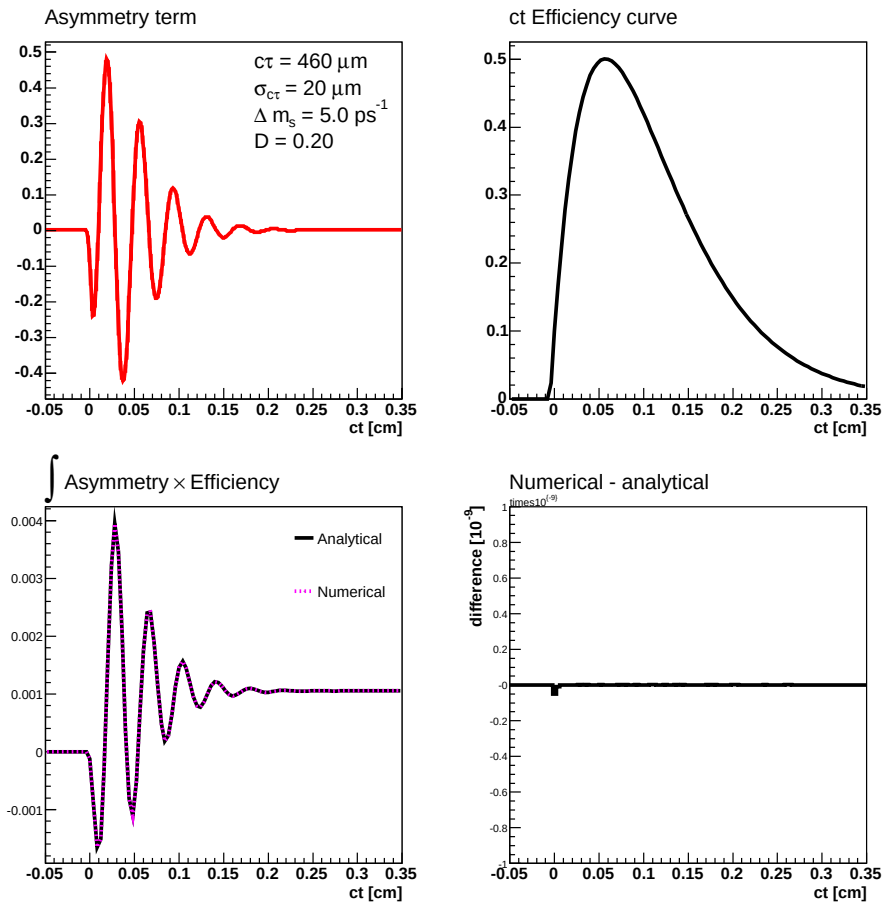


Figure C.2: Comparison of the implemented analytical solution of the sum of integrals $N_{10} + N_{11} + N_{12}$, this corresponds to the normalization of the function $f_2(t_i, \xi_i, \sigma_{t,i})$, to a numerical integration. The top left panel shows $f_2(t_i, \xi_i, \sigma_{t,i})$ for some typical values of $c\tau$, σ_t, Δm, D and $\xi = 1$, the top right panel shows the efficiency curve that is used, the bottom left panel shows the analytical and numerical integrals $\int_{-0.05}^{x} f_2(t_i, \xi_i, \sigma_{t,i})$ superimposed, where the x–axis denotes the upper limit used for the integration, and the bottom right panel the difference of the two.

Appendix D

Comparisons between Data and Simulated Events

D.1 Overview

This section summarizes the comparisons of simulated events to data for all variables that have been used for the training of a neural network throughout the scope of this note. First (section D.2) a comparison between data and simulated events obtained after a cut based selection is presented for the decay $B^0 \to D\pi, D \to K\pi\pi$. The comparisons demonstrates that the agreement between data and simulated events is generally good and the simulated sample can be used to provide training patterns for signal events in the network training. In the following section D.3 the correlations between the variables in data and simulated events are compared, this check shows that the correlations the neural network is going to learn from the simulation are the same as in data. Finally section D.4 demonstrates that the agreement between data and simulated events is still valid when applying a neural network selection on both data and simulated events. This is important when deriving the lifetime efficiency curve from the lifetime distribution of simulated signal events (see section 7.2.2). All checks have been performed for the decay $B_s \to D_s\pi, D_s \to \phi\pi, \phi \to KK$ as well and show a similarly good description of signal events in data by events obtained from a simulation.

For the preparation of the plots in each section the signal region is defined as reconstructed B^0 mass $\pm 2\sigma$ where σ denotes the width of the narrower Gaussian functions in the signal peak obtained from the fit in figure 6.9 (top). This covers the mass range [5.2419, 5.3163] GeV. The sideband region is chosen to be the range from [5.4, 5.6] GeV. Before subtracting, the sideband distributions are scaled to the number of events expected for the combinatorial background contribution in the signal region.

The sample of simulated events is scaled to the number of true signal events (excluding contributions from the Cabibbo–suppressed mode $B^0 \to DK$). Each plot shows the scaled distribution for simulated events (grey), the sideband subtracted signal from data (black markers), both normalized to unit area, and for reference the distribution of the variable in the upper sideband (blue markers).

D.2 Comparison of Distributions after Cuts

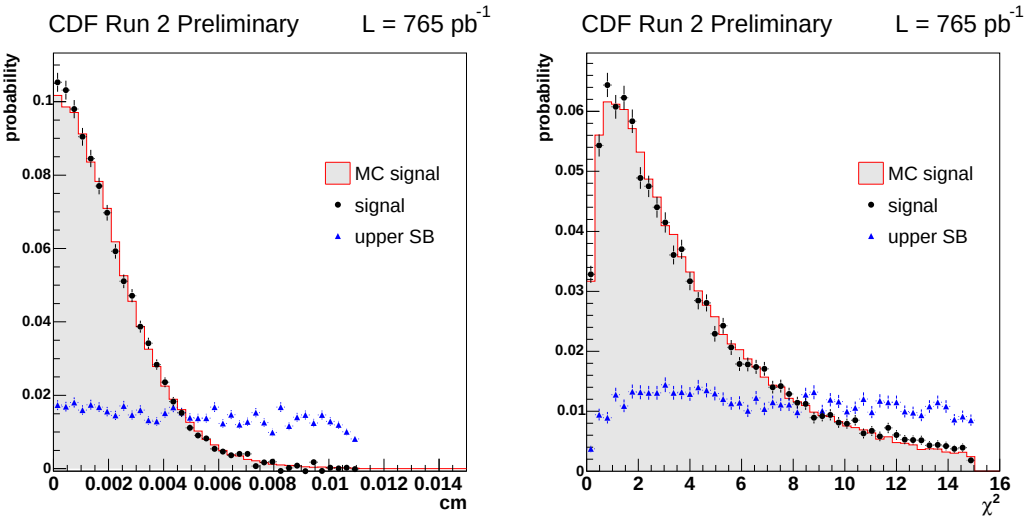


Figure D.1: Comparison of data and simulated events of the impact parameter of $|d_0(B)|$ (left) and the two–dimensional χ^2 of the vertex fit of the B^0 (right).

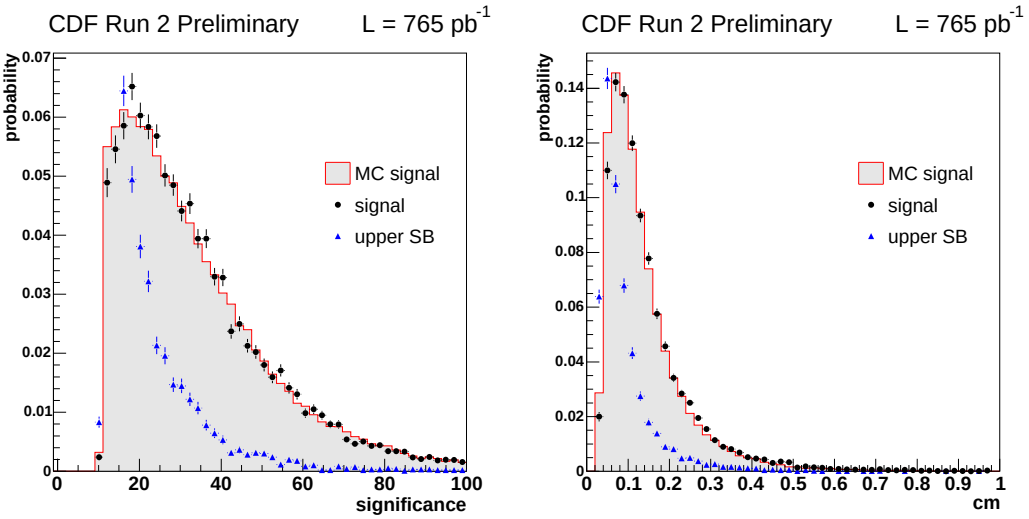


Figure D.2: Comparison of data and simulated events of the decay length significance $L_{xy}/\sigma_{L_{xy}}(B)$ of the B^0 (left) and the distribution of the decay length $L_{xy}(B)$ of the B^0 (right).

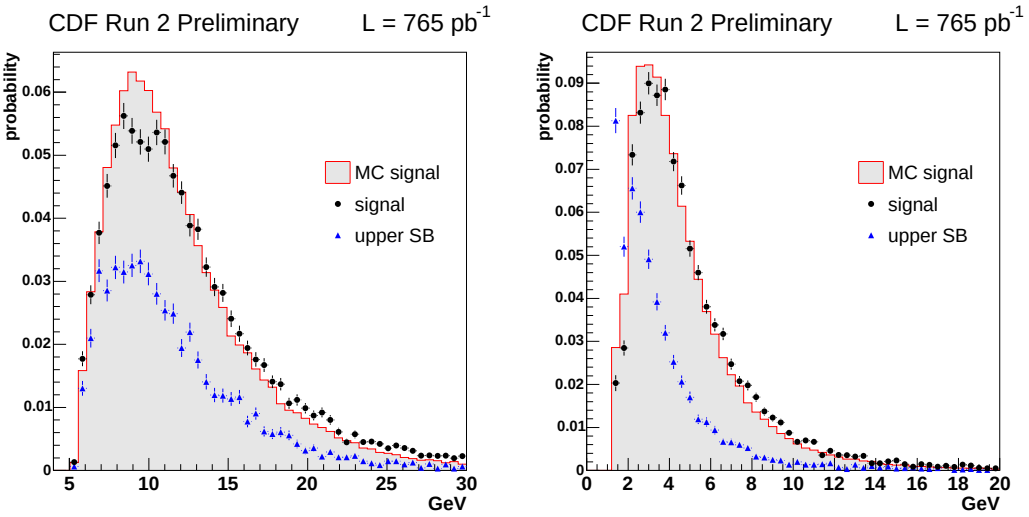


Figure D.3: Comparison of data and simulated events of the p_t of the B^0 meson and the p_t of the pion originating from the decay $B^0 \to D\pi$ (right).

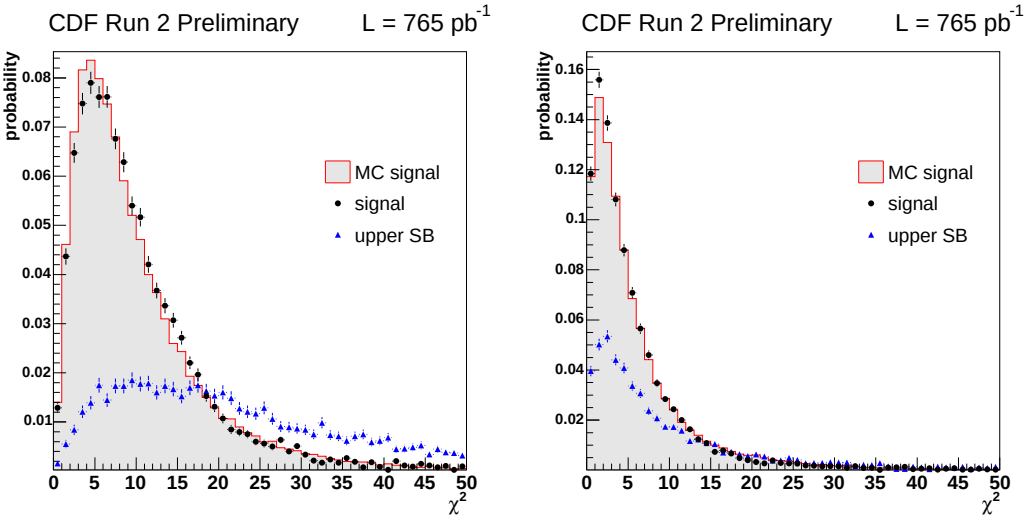


Figure D.4: Comparison of data and simulated events of three–dimensional χ^2 of the vertex fit of the decay $B^0 \to D\pi, D \to K\pi\pi$ (left) and the decay $D \to K\pi\pi$ (right).

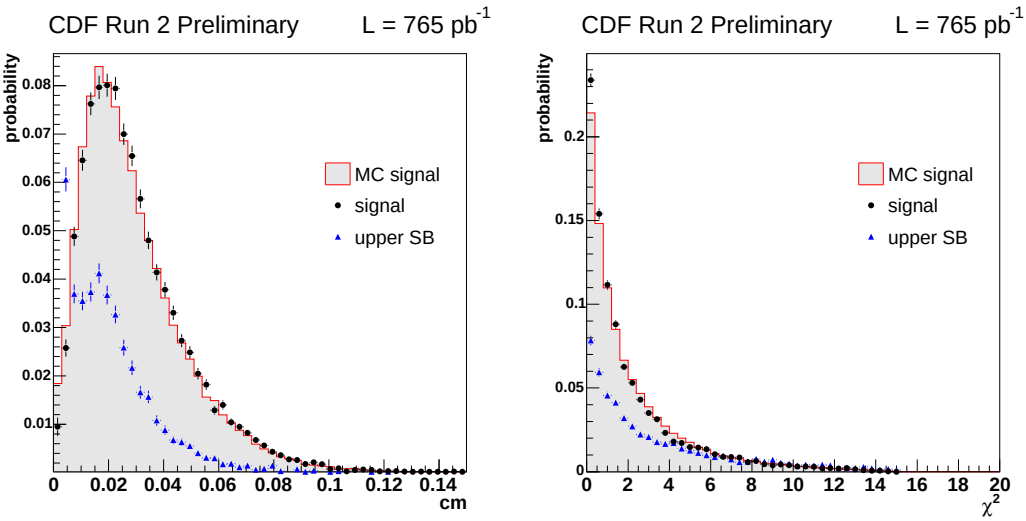


Figure D.5: Comparison of data and simulated events of the impact parameter (left) and the two–dimensional χ^2 of the vertex fit of the D from the decay $B^0 \to D\pi$ (right).

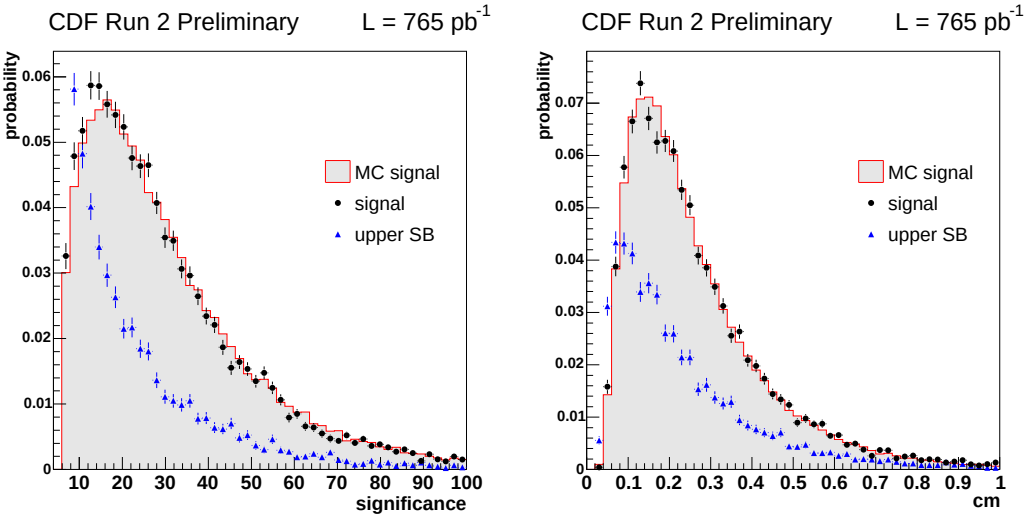


Figure D.6: Comparison of data and simulated events of the decay length significance $L_{xy}/\sigma_{L_{xy}}(B)$ (left) and the distribution of the decay length $L_{xy}(D)$ of the D from the decay $B^0 \to D\pi$ (right).

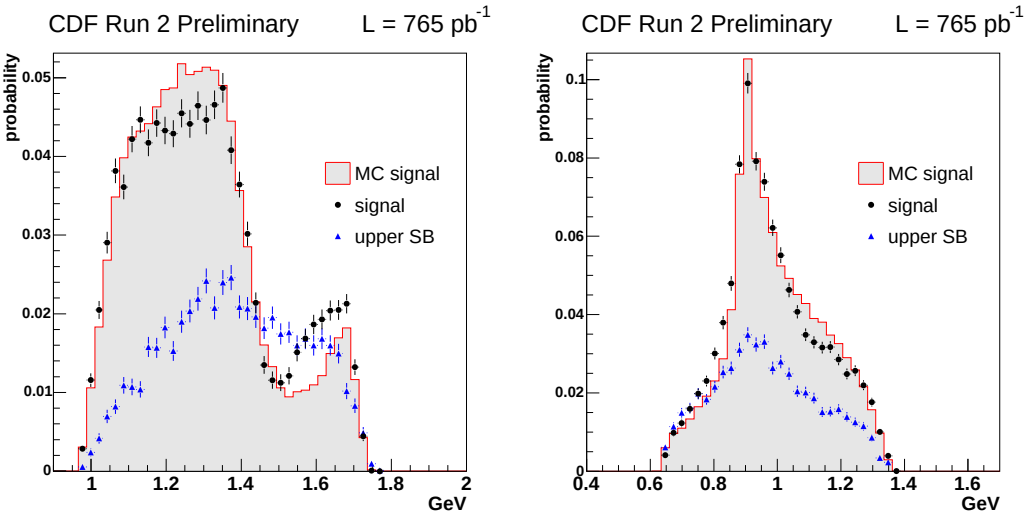


Figure D.7: Comparison of data and simulated events of the invariant masses of the subsystems m_{K,π^1_D} and m_{K,π^2_D} from the decay $D \to K\pi\pi$. The two masses are ordered, the left panel shows the higher mass $\max(m_{K,\pi^1_D}, m_{K,\pi^2_D})$, the smaller mass $\min(m_{K,\pi^1_D}, m_{K,\pi^2_D})$ is displayed in the right panel.

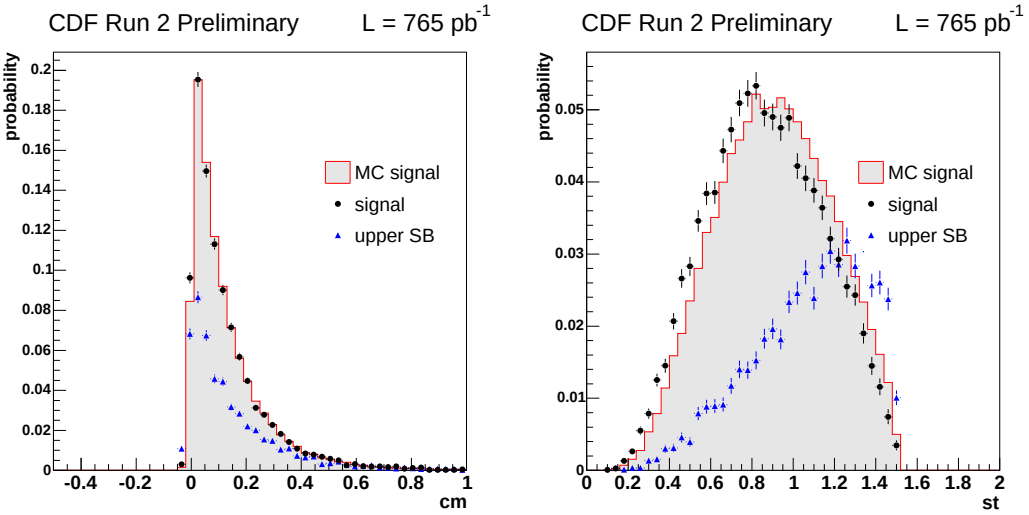


Figure D.8: Comparison of data and simulated events of the decay length of the D from the decay $B^0 \to D\pi$ with respect to the B^0 (left) and the opening angle $\Delta R(D, \pi_B)$ between D and π from the $B^0 \to D\pi$ decay (right).

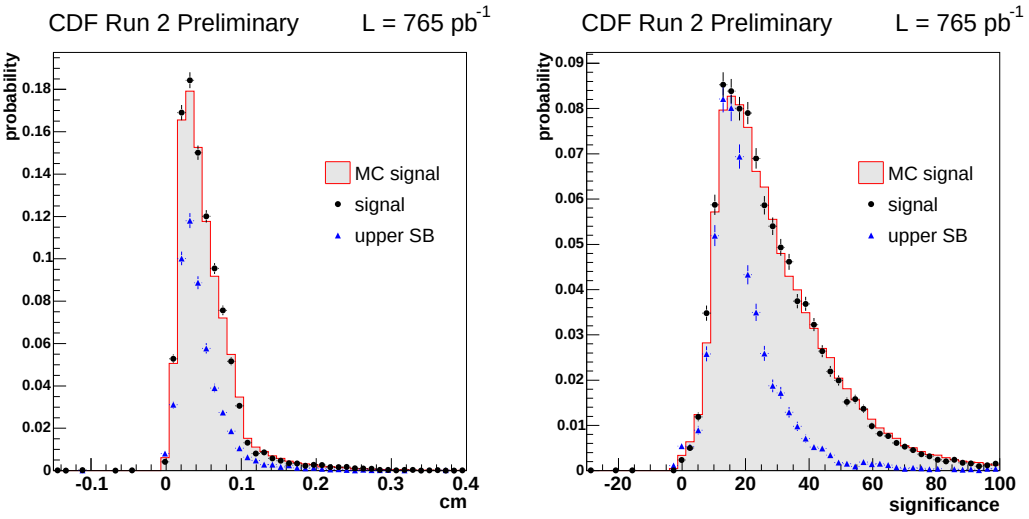


Figure D.9: Comparison of data and simulated events of the the lifetime signed impact parameter d_0^{lts} (left) and the lifetime signed impact parameter significance d_0^{lts}/σ_{d_0} of the π from the $B^0 \to D\pi$ decay (right).

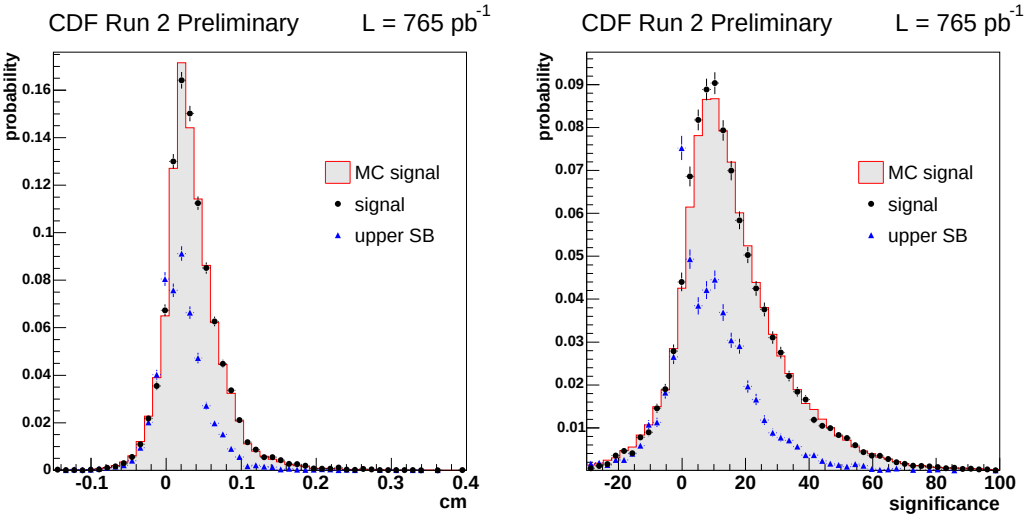


Figure D.10: Comparison of data and simulated events of the the lifetime signed impact parameter d_0^{lts} (left) and the lifetime signed impact parameter significance d_0^{lts}/σ_{d_0} of the K from the $D \to K\pi\pi$ decay (right).

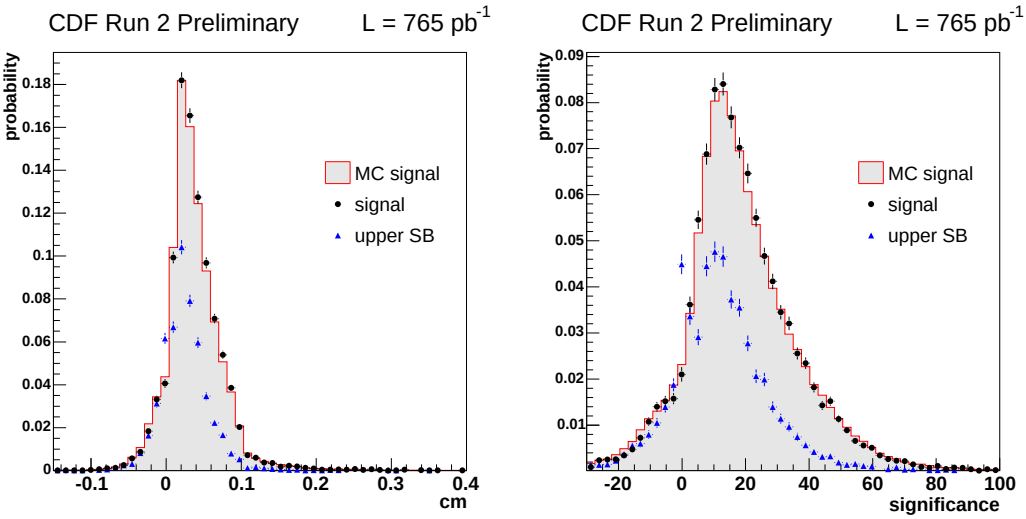


Figure D.11: Comparison of data and simulated events of the the lifetime signed impact parameter d_0^{lts} (left) and the lifetime signed impact parameter significance d_0^{lts}/σ_{d_0} of the first π from the $D \to K\pi\pi$ decay (right).

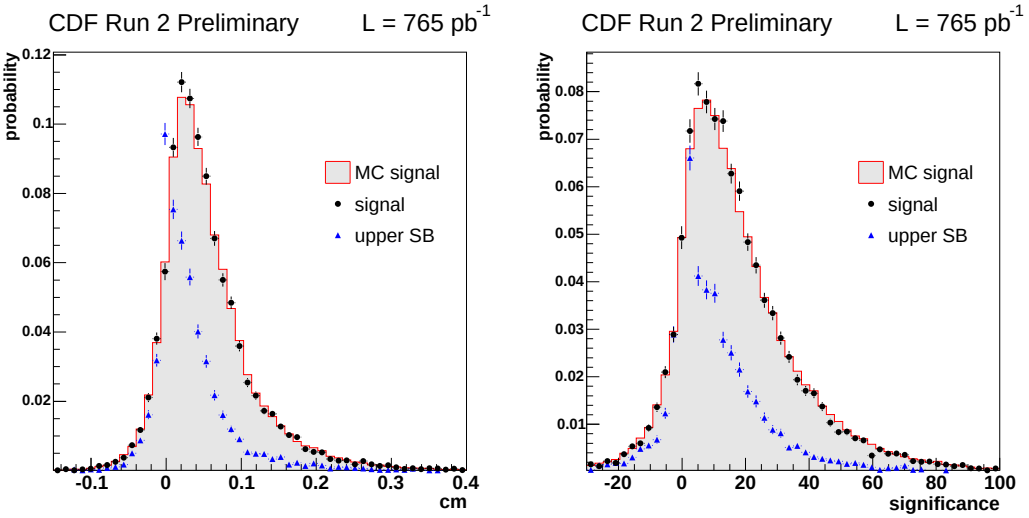


Figure D.12: Comparison of data and simulated events of the lifetime signed impact parameter d_0^{lts} (left) and the lifetime signed impact parameter significance d_0^{lts}/σ_{d_0} of the second π from the $D \to K\pi\pi$ decay (right).

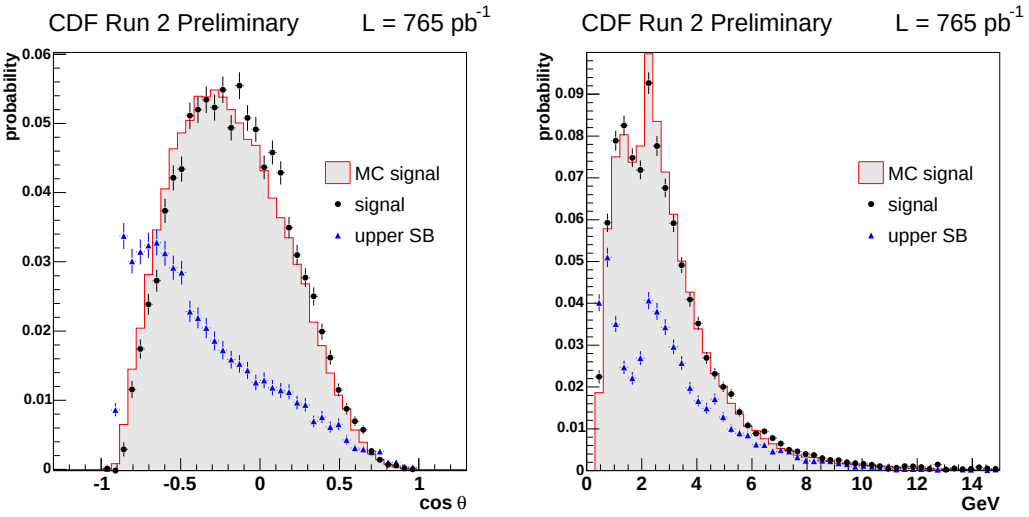


Figure D.13: Comparison of data and simulated events of the angle $\sphericalangle(\vec{p}(B), \vec{p}(\pi_B))_{\mathrm{cms}B}$ between $\vec{p}(\pi_B)$ and $\vec{p}(B)$ in the B rest frame (left) and the p_t of the K from the decay $D \to K\pi\pi$ (right).

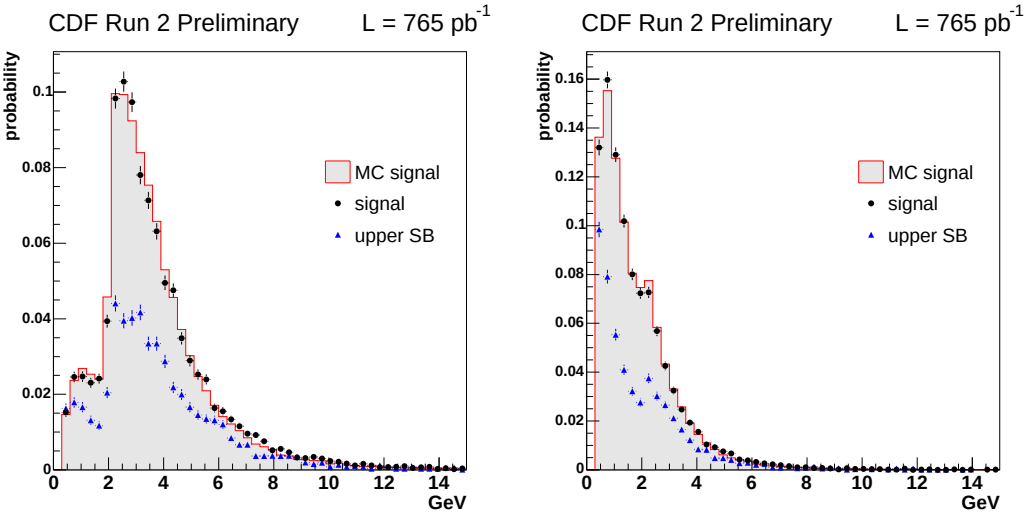


Figure D.14: Comparison of data and simulated events of the transverse momenta p_t of the first π (left) and the second π from the decay $D \to K\pi\pi$ (right).

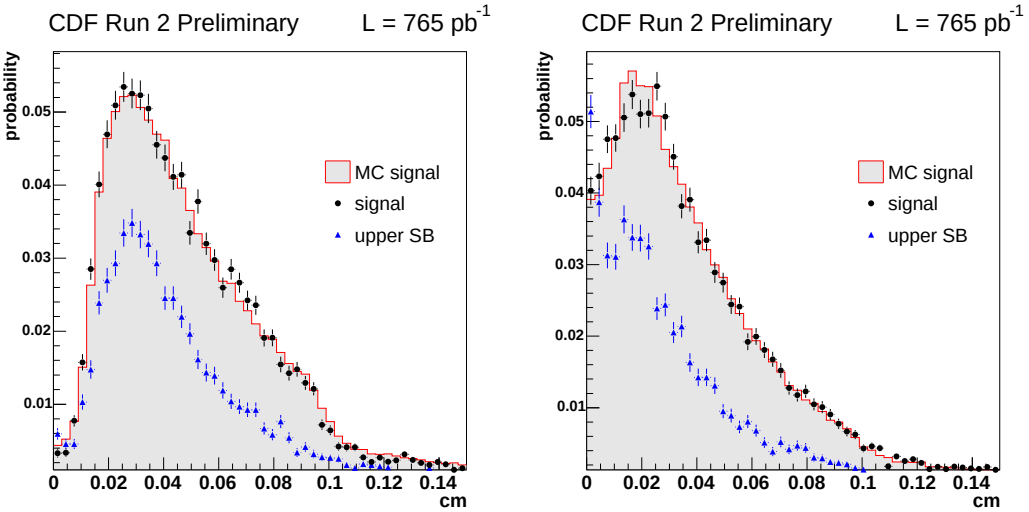


Figure D.15: Comparison of data and simulated events of the impact parameters $|d_0|$ of the π_B (left) and the K from the decay $D \to K\pi\pi$ (right).

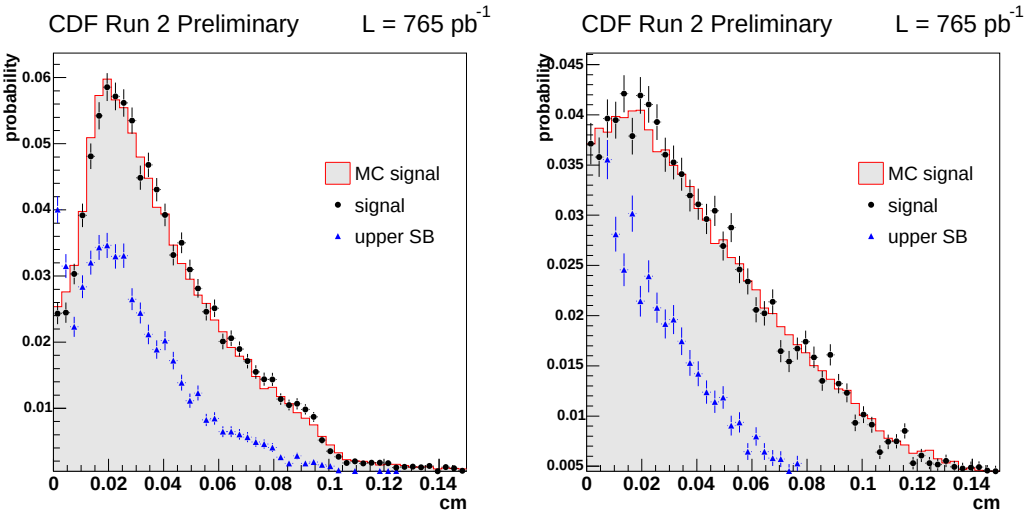


Figure D.16: Comparison of data and simulated events of the impact parameters $|d_0|$ of the first π (left) and the second π from the decay $D \to K\pi\pi$ (right).

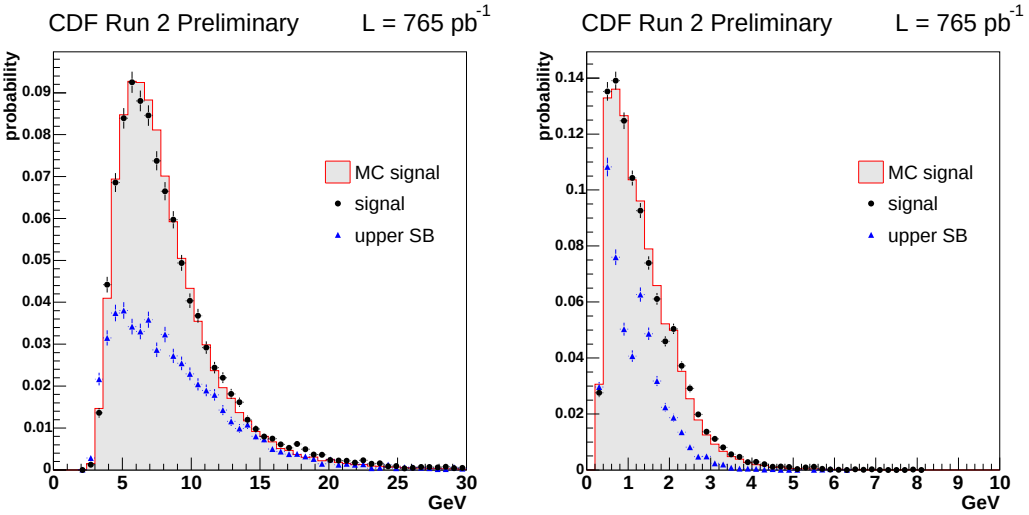


Figure D.17: Comparison of data and simulated events of the p_t of the D meson from the decay $B^0 \to D\pi$ (left) and the minimum p_t of the four stable particles from the decay chain $B^0 \to D\pi, D \to K\pi\pi$ (right).

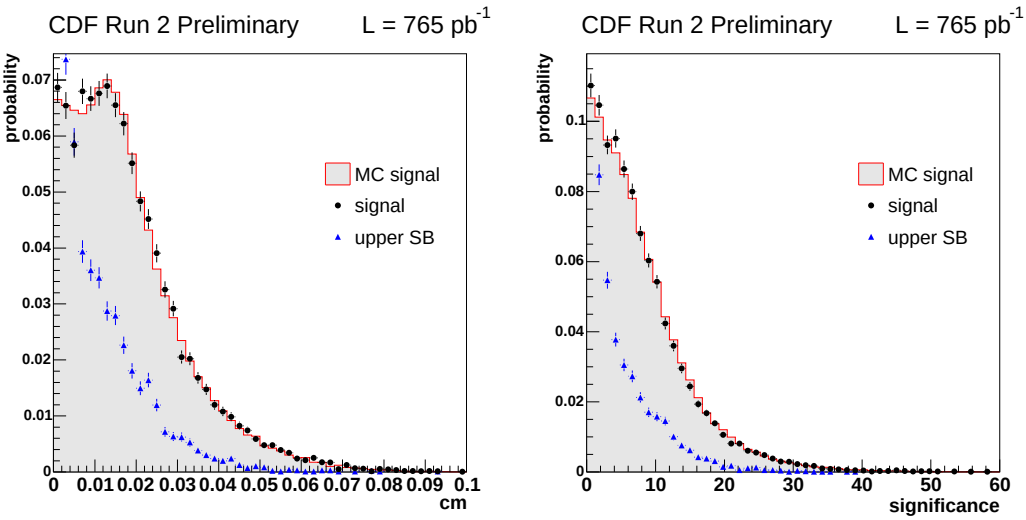


Figure D.18: Comparison of data and simulated events of the minimum impact parameter $|d_0|$ and the minimum impact parameter significance $|d_0|/\sigma_{d_0}$ of the four stable particles from the decay chain $B^0 \to D\pi, D \to K\pi\pi$.

D.3 Comparison of Correlations in Data and Simulated Events

This section summarizes the data/simulated events comparison of the *correlations* between the training variables. As the main benefit from using a network instead of applying cuts on individual variables arises from taking correlations into account, it's important to check that the simulated events simulation properly reproduces the correlations that are observed in real data. The plots in this section are prepared using the same sideband subtraction procedure as in the previous section. For each pair of variables 2–dimensional histograms are prepared for the signal range in data, the upper sideband region from 5.4 to 5.6 GeV in data and the signal from simulated events. The distribution from the upper sideband is then subtracted from the distribution in data. For both data and simulated events then profile plots are created. The following pages summarize the results of this comparison for most of the variables.

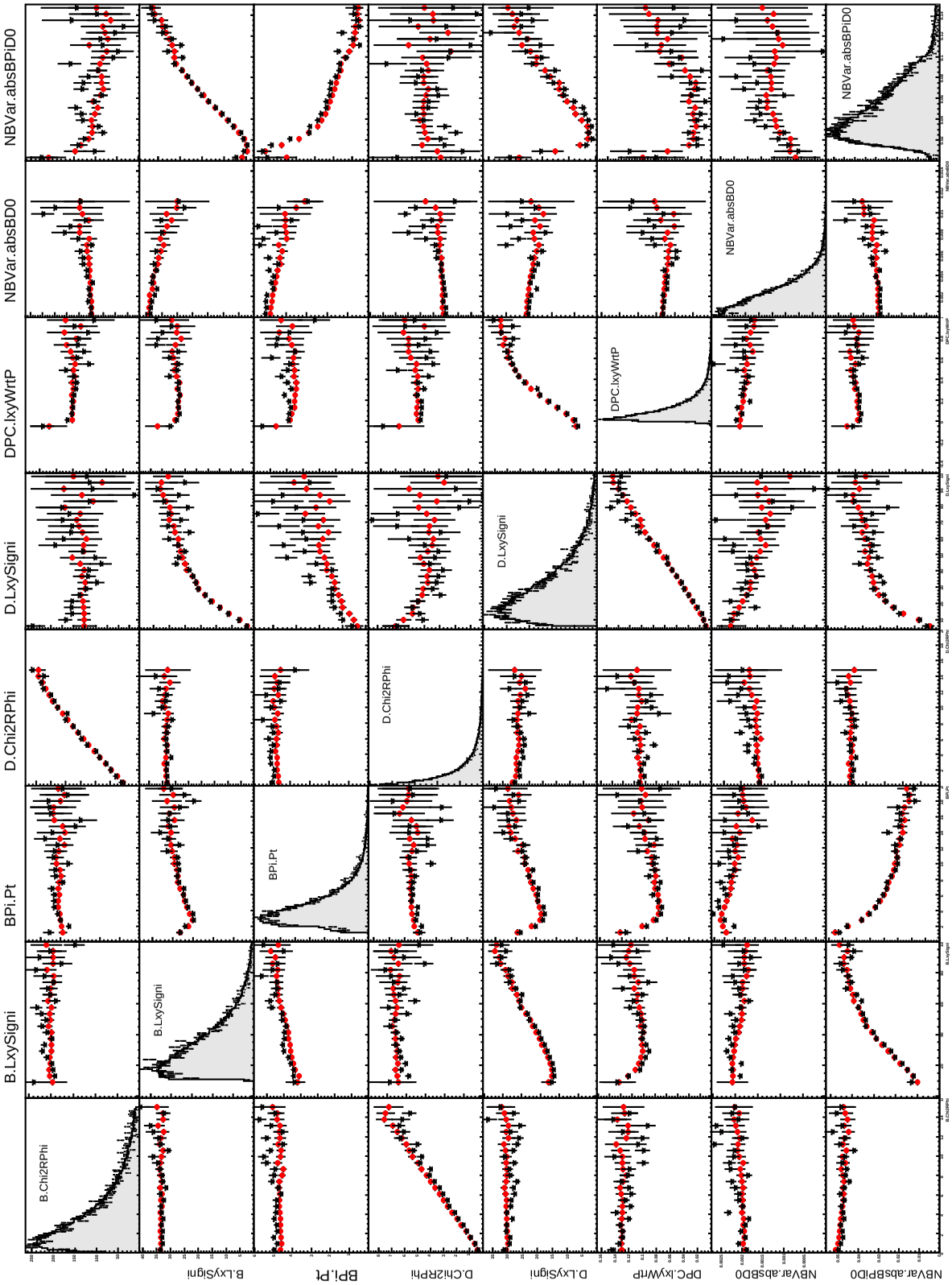


Figure D.19: Comparison of data and simulated events of correlations between variables. Red denotes the simulated events simulation, black denotes data.

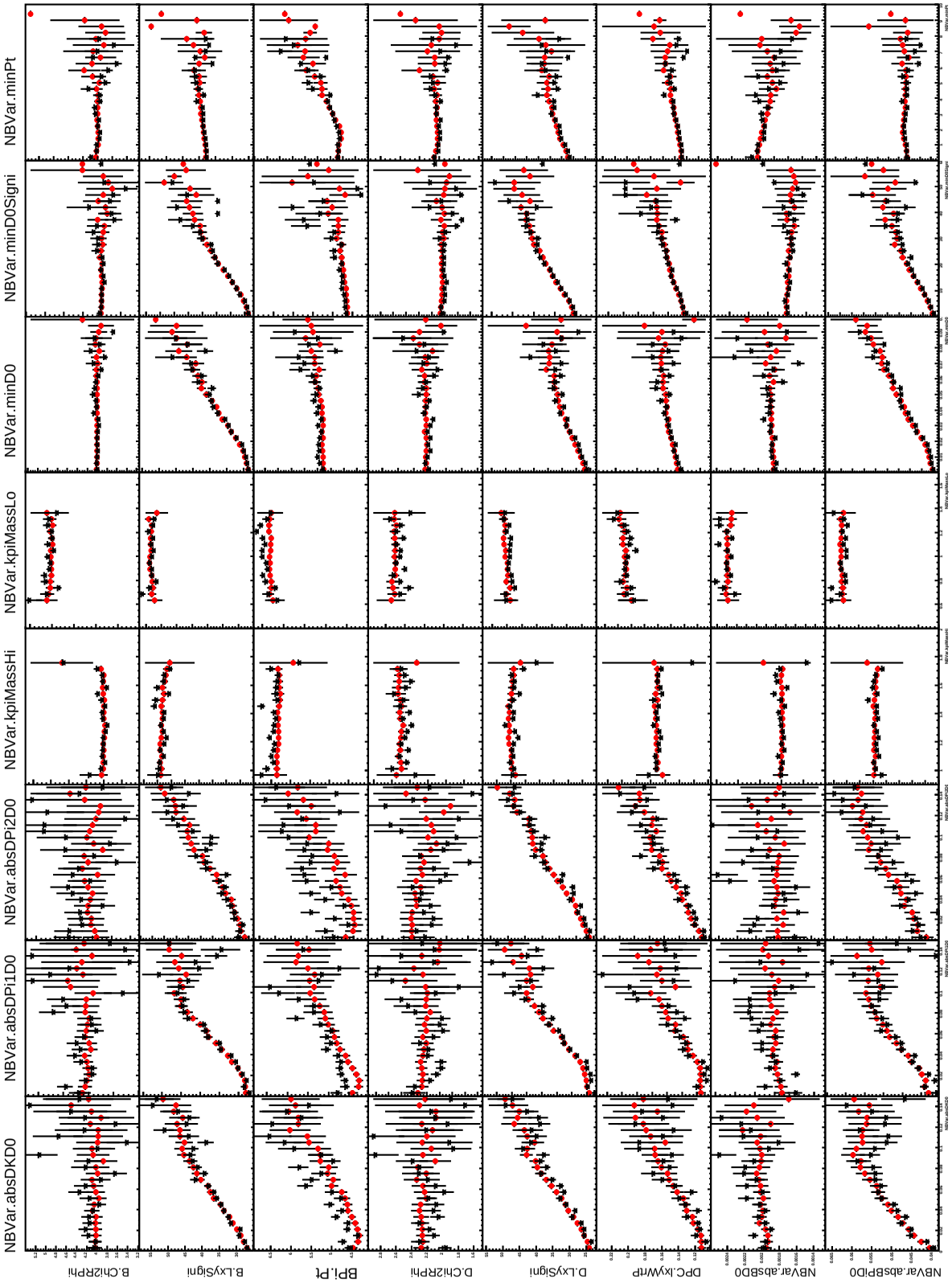


Figure D.20: Comparison of data and simulated events of correlations between variables. Red denotes the simulated events simulation, black denotes data.

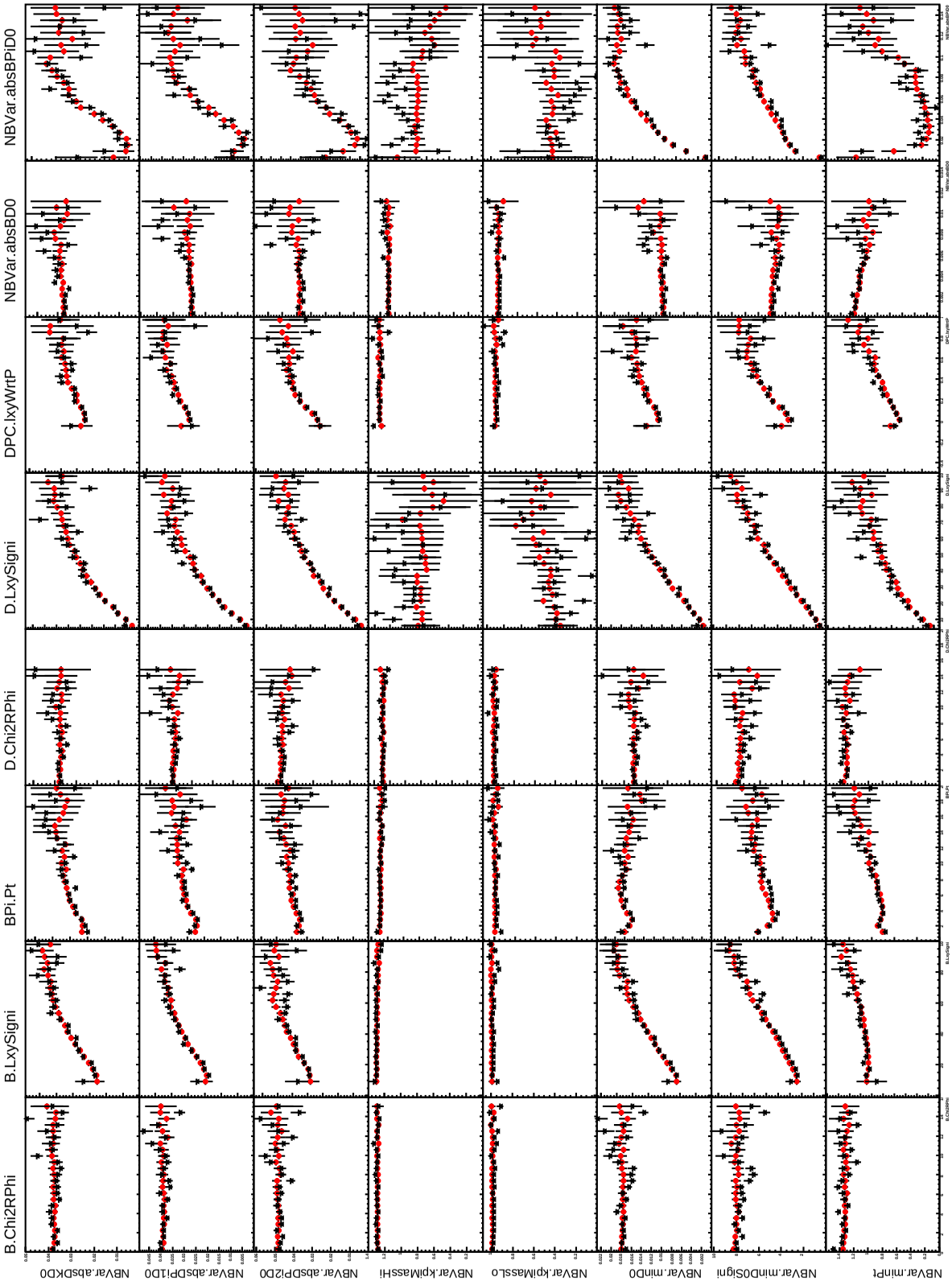


Figure D.21: Comparison of data and simulated events of correlations between variables. Red denotes the simulated events simulation, black denotes data.

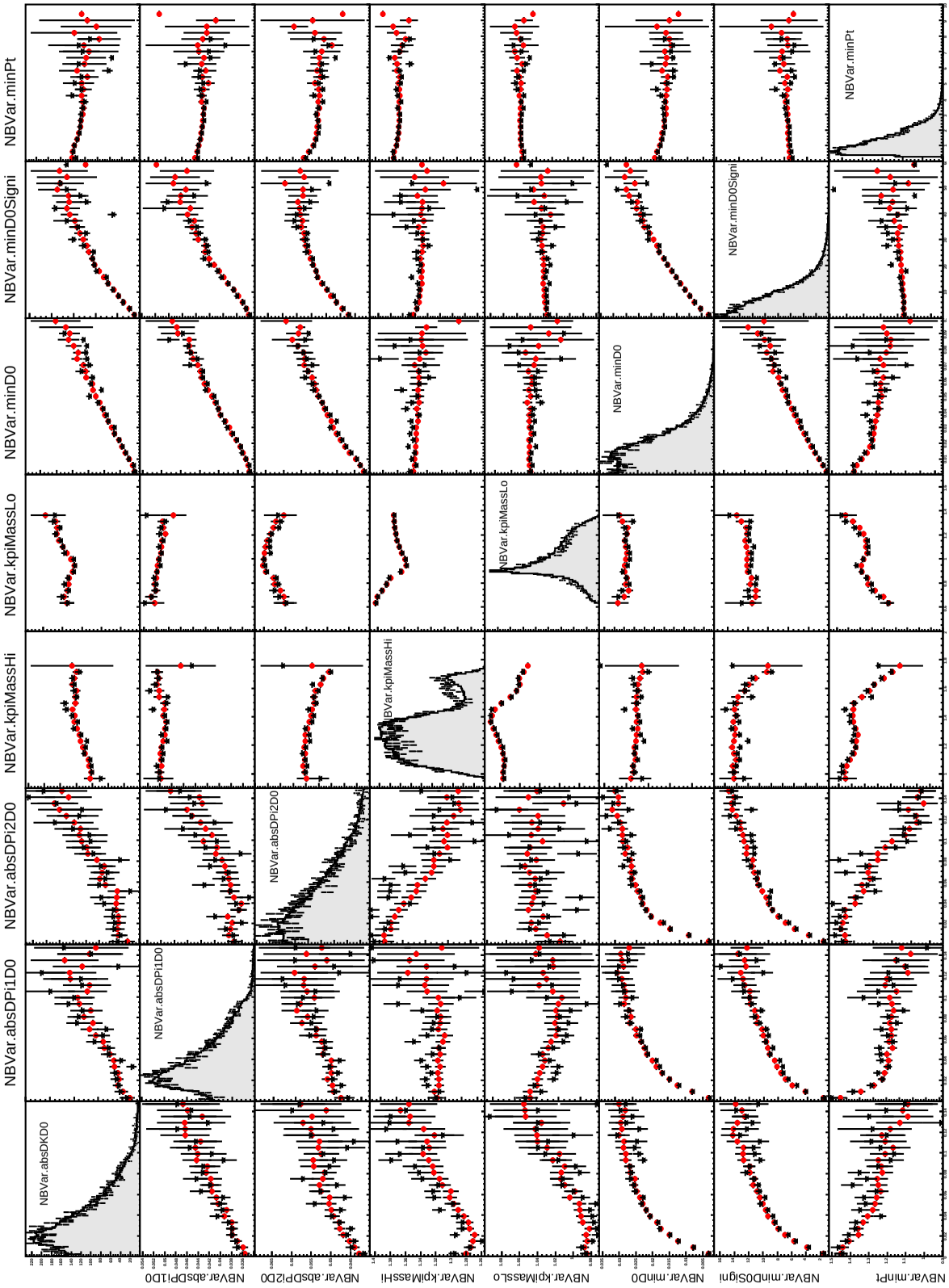


Figure D.22: Comparison of data and simulated events of correlations between variables. Red denotes the simulated events simulation, black denotes data.

D.4 Comparison of Data and Simulated Events after a Network Selection

This section summarizes the comparisons of simulated events to data for all variables that have been used for the training *after a neural network selection, applying the network cut that optimized* $\frac{S}{\sqrt{S+B}}$. All distributions are prepared in analogy to those already shown in appendix D.2. Each plot shows the scaled distribution from simulated events (grey), the sideband subtracted signal from data (black markers), both normalized to unit area, and for reference the distribution of the variable in the upper sideband (blue markers). For the compilation here distributions of the variables from the reduced set of training variables compiled in table 6.2 in section 6.5 is displayed.

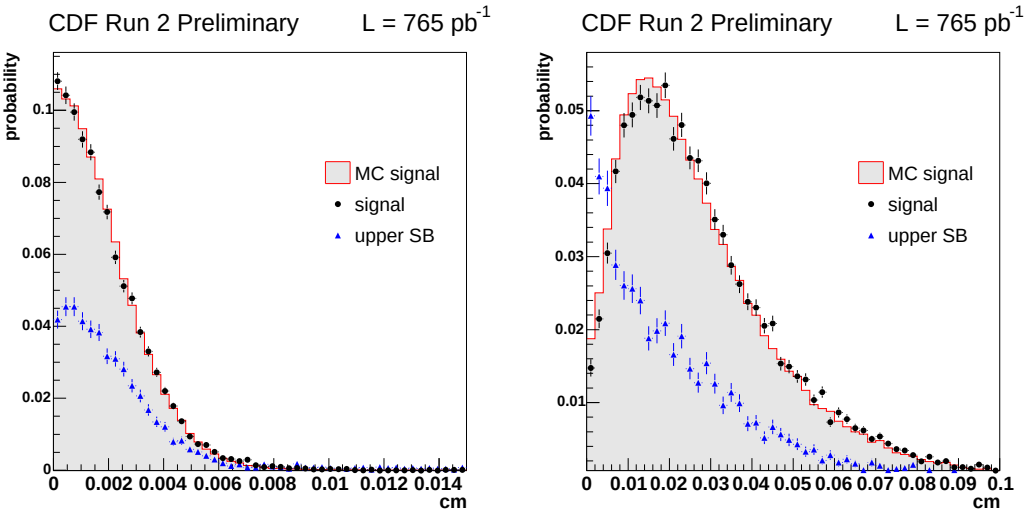


Figure D.23: Comparison of data and simulated events of the impact parameter $|d_0(B)|$ of the B (left) and the impact parameter $|d_0(D)|$ of the D (right).

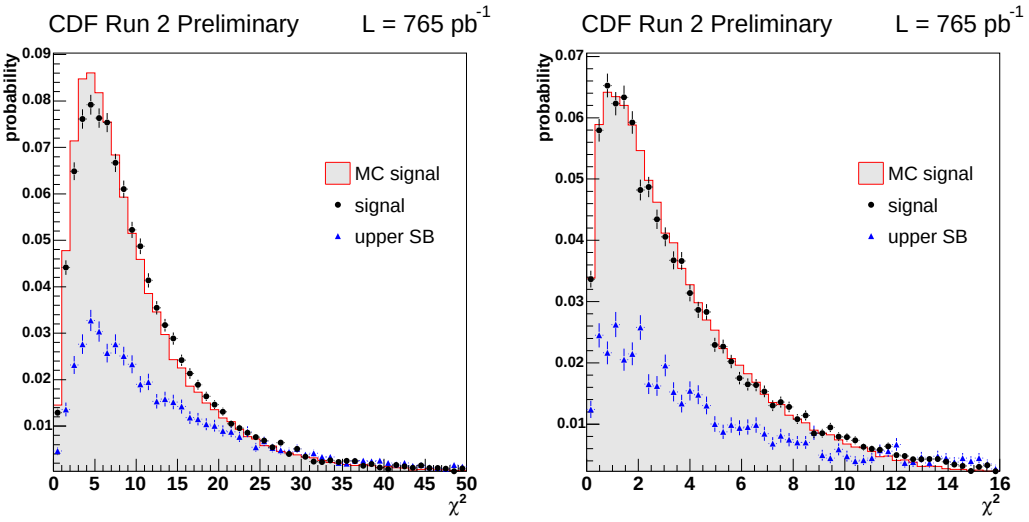


Figure D.24: Comparison of data and simulated events of the three– (left) and two–dimensional (right) χ^2 of the vertex fit of t he B^0 (right).

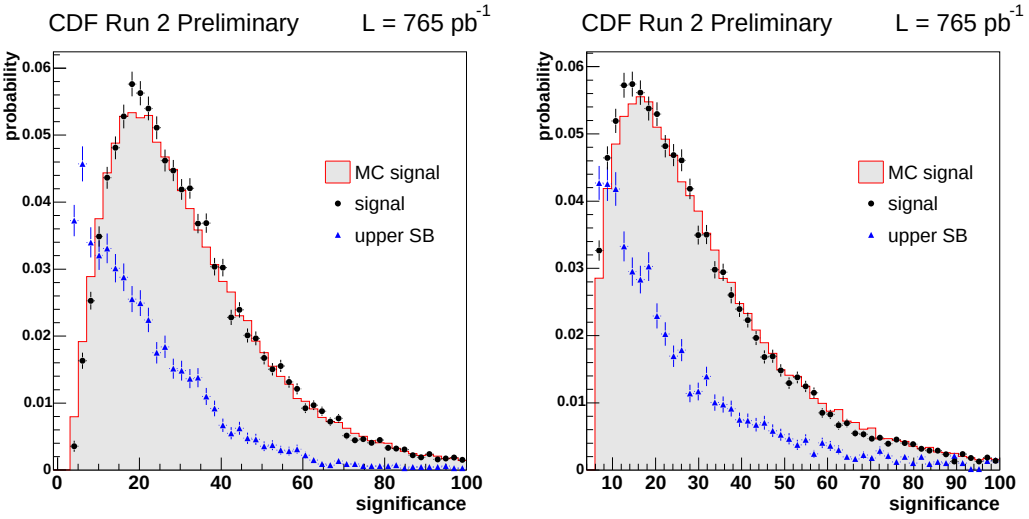


Figure D.25: Comparison of data and simulated events of the decay length significance $L_{xy}/\sigma_{L_{xy}}(B)$ of the B^0 (left) and the decay length significance $L_{xy}/\sigma_{L_{xy}}(D)$ of the D (right).

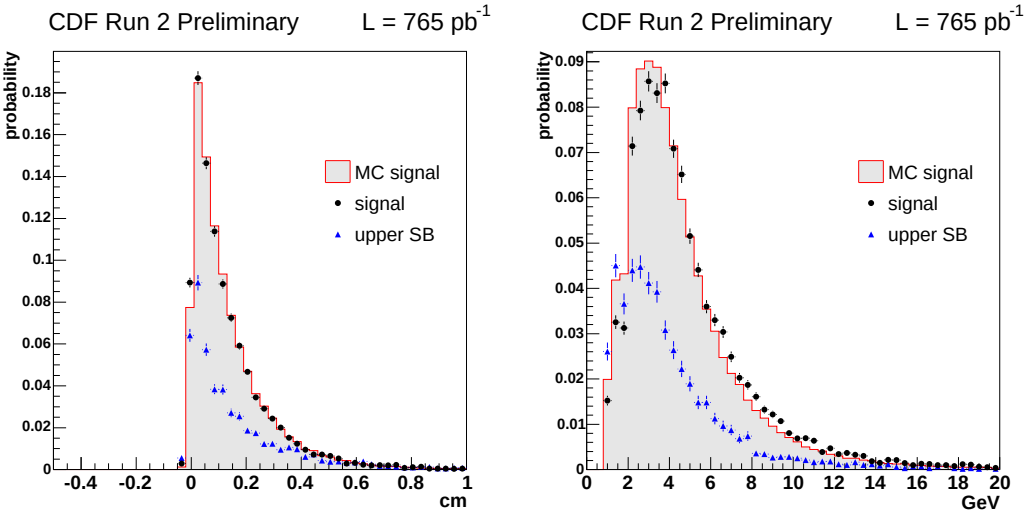


Figure D.26: Comparison of data and simulated events of the L_{xy} of the D with respect to the B vertex and the p_t of the pion originating from the decay $B^0 \to D\pi$ (right).

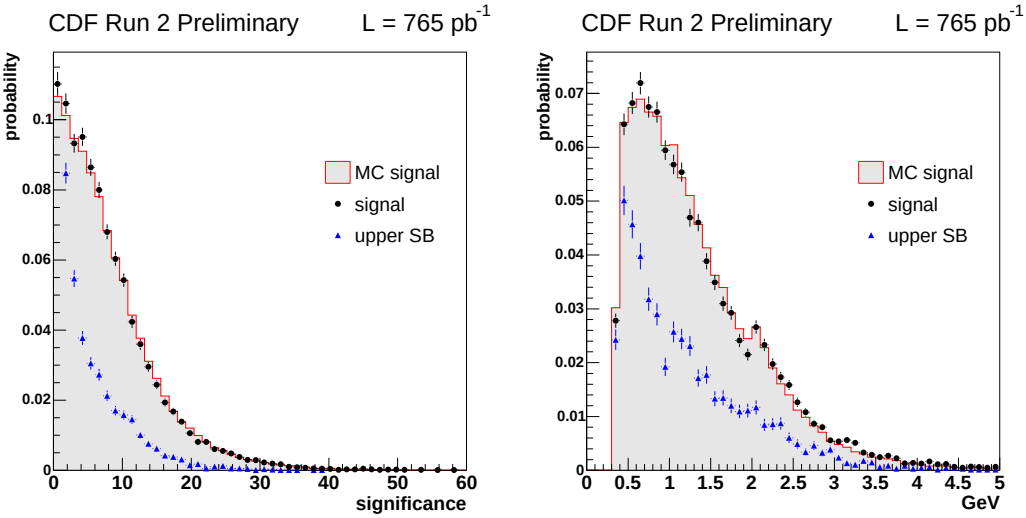


Figure D.27: Comparison of data and simulated events of the minimum impact parameter significance $|d_0|/\sigma_{d_0}$ (left) and the minimum p_t of the four stable particles from the decay chain $B^0 \to D\pi, D \to K\pi\pi$ (right).

Appendix E

Compilation of Fit Templates

This appendix shows an exemplaric compilation of background templates that were used for fits of the B meson lifetime. All templates are rederived for different network cuts by applying to the corresponding network cuts on the sample used to fit the template. The templates shown here enter the lifetime fits in figure 7.4. The function used to describe the lifetime templates is in every case the sum of two exponential functions convolved with two Gaussian functions (cf. 7.2):

$$P_B(t) = N_B \left(f_1 \, e^{-\frac{t}{\tau_1}} \otimes \mathcal{G}(t - t', \mu_1, \sigma_1) \; + \; (1 - f_1) \, e^{-\frac{t}{\tau_2}} \otimes \mathcal{G}(t - t', \mu_2, \sigma_2) \right)$$

All parameters of the two exponential functions and the two Gaussian functions are released in the template fits, in total there are eight free parameters for each template fit. The parameters of the upper sideband template fit only enter the lifetime fit as start values, except for the overall normalization, which is fixed to one, and σ_1/σ_2, which is fixed to the value obtained after the template fit, all parameters are floating in the unbinned maximum likelihood fit. The parameters describing the lifetime distribution of the decays $\Lambda_b \to \Lambda_c \pi$ and $B_s \to D_s \pi$ are kept fixed during the unbinned maximum likelihood fit.

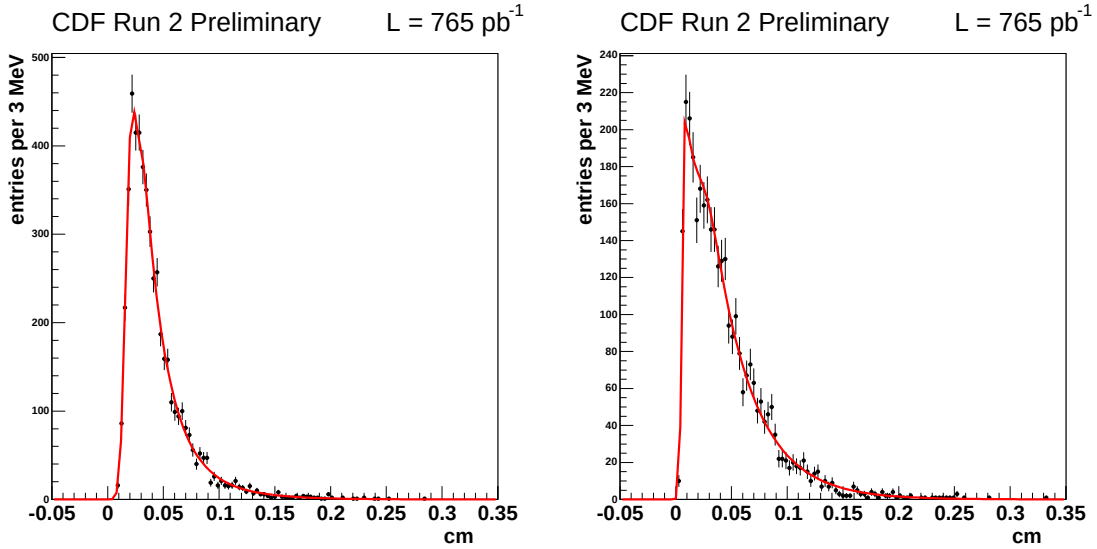


Figure E.1: Templates describing the contribution of combinatorial background from 5.4 to 5.6 GeV in the lifetime distribution. The distribution displayed in the left panel was obtained after a cut based selection, the distribution in the panel on the right hand side results from a neural network selection. The parameters of the fitted function give the start values for the background contributions for the fits displayed in figure 7.4.

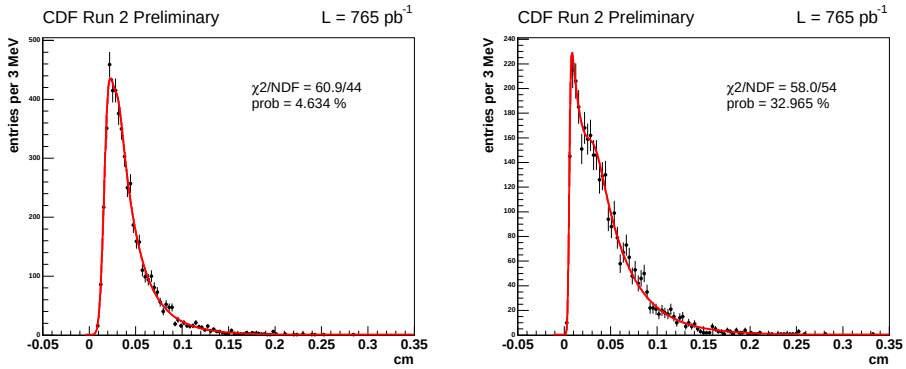


Figure E.2: Projection of the fit results from the unbinned maximum likelihood fit of the background contribution of combinatorial background superimposed on the lifetime distribution obtained for combinatorial background in the range from 5.4 to 5.6 GeV. The distribution displayed in the left panel is obtained after a cut based selection, the distribution in the panel on the right hand side results from a network based selection. The parameters of the superimposed functions result from the fit displayed in figure 7.4.

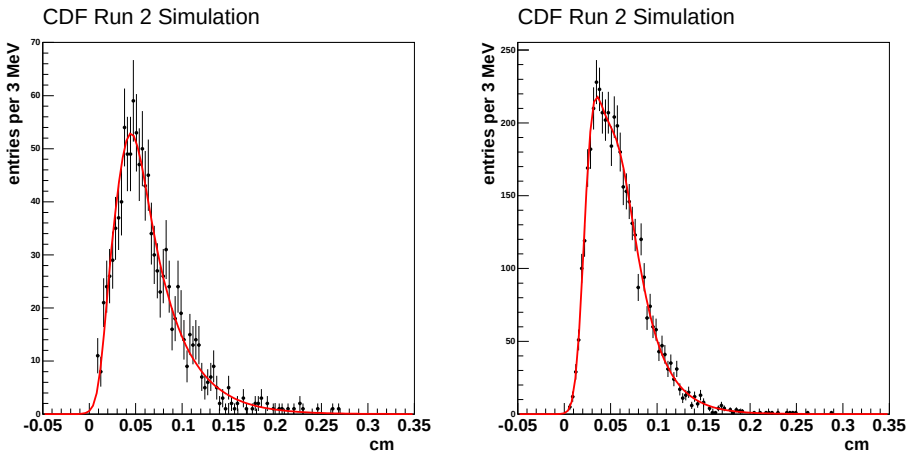


Figure E.3: Templates obtained after a neural network selection describing the contributions of the decays $\Lambda_b \to \Lambda_c \pi$ and $B_s \to D_s \pi$ reconstructed as $B^0 \to D\pi$ to the lifetime distribution. The parameters of the fitted functions are used in the fit shown in the right panel of figure 7.4.

Bibliography

[1] S. L. Glashow. Nucl. Phys., 22:579, 1961.

[2] S. Weinberg. Phys. Rev. Lett., 19:1264, 1967.

[3] A. Salam. in *Elementary Particle Theory*, 1968.

[4] D. J. Gross and F. Wilczek. Phys. Rev., D8:3633–3652, 1973.

[5] M. Kobayashi and T. Maskawa. Prog. Theor. Phys., 49:652, 1973.

[6] N. Cabibbo. Phys. Rev. Lett., 10:531–532, 1963.

[7] E.E. Lande et al. Phys. Rev., 103:1901, 1956.

[8] C. Albajar et al. (UA1 Collaboration). Phys. Lett., B186:247, 1987.

[9] H. Albrecht et al. Phys. Lett., B192:245, 1987.

[10] The Heavy Flavor Averaging Group. *Summary of Averages as of End of 2005.* `hep-ex/0603003`.

[11] The CDF Collaboration. *Measurement of the $B_s - \bar{B}_s$ Oscillation frequency*, 2006. CDF Public Note 8204, available at `http://www-cdf.fnal.gov/physics/new/bottom/bottom.html`.

[12] H. G. Moser and A. Roussarie. *Mathematical Methods for $B_s - \bar{B}_s$ oscillation analyses.* NIM, A384:491–505, 1997.

[13] M. Feindt, S. Menzemer, and K. Rinnert. *TrackingKal - A Tracking and Alignment Software Package for the CDFII Silicon Detector*, 2003. CDF/THESIS/TRACKING/PUBLIC/5968.

[14] Torbjorn Sjostrand, Leif Lonnblad, and Stephen Mrenna. Pythia 6.2: Physics and manual. 2001. LU TP 01-21, `hep-ph/0108264`.

[15] S. et al. Eidelman. Review of Particle Physics. Physics Letters B, 592:1+, 2004.

[16] C. D. Anderson. Physical Review, 43:491, 1933.

[17] C. P. Burgess and G. D. Moore. *A Primer on the Standard Model*, 2003. Lecture notes, Mc Gill University.

[18] L. Wolfenstein. Phys. Rev. Lett., 51:1945, 1983.

[19] The CKMFitter Group. *CP violation and the CKM matrix*. Eur.Phys.J., C41:1–131, 2005.

[20] B. Aubert et al. (The BaBar Collaboration). Phys. Rev. Lett., 90:181801, 2003.

[21] S.B. Athar et al. (The CLEO Collaboration). Phys. Rev., D68:072003, 2003.

[22] The Heavy Flavour Averaging Group. Summer 2003 and winter 2004 averages, 2003, 2004. `http://slac.stanford.edu/xorg/hfag/`.

[23] M. Ciuchini, E. Franco, V. Lubicz, F. Mescia, and C. Tarantino. *Lifetime differences and CP violation parameters of neutral B mesons at the next-to-leading order in QCD*. JHEP, 08:031, 2003.

[24] M. A. Mazzoni. *B mixing, lifetimes, lifetime differences and CP / T / CPT violation at the B factories*. Prepared for 16th Conference on High Energy Physics (IFAE 2004), Turin, Italy, 14-16 Apr 2004.

[25] A. J. Buras et al. Nucl. Phys., B345:369, 1984.

[26] A. Kronfeld. `hep-lat/0310063v1`.

[27] A. J. Buras et al. Nucl. Phys., B347:491, 1990.

[28] A. Gray et al. (HPQCD Collaboration). Phys. Rev. Lett., 95:212001, 2005.

[29] The CDF II Collaboration. *The CDF II Detector Technical Design Report*, 1996. FERMILAB-Pub-96/390-E.

[30] A. Mukherjee and R. Wagner. *COT Central Outer Tracker*, 2003. CDF/PUB/TRACKING/PUBLIC/6267.

[31] G. Gomez for the TOF group. *The CDF II Time–of–Flight–detector*, 2003. CDF Note 6258.

[32] D. Lucchesi. *The CDF II Seconday Vertex Trigger*, 2002. Proceedings of Beauty 2002 and CDF Note 6107.

[33] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*, 1998. B.G.Teubner, Stuttgart.

[34] M. Feindt. *A Neural Bayesian Estimator for Conditional Probability Densities*, 2004. `physics/0402093`.

[35] The late Rev. Mr. T. Bayes. *An Essay towards solving a Problem in the Doctrine of Chances*. Philosophical Transactions of the Royal Society of London, 53:370–418, 1763.

[36] Boris Y Iyutin. Run–averaged prescale factors. `http://mit1.fnal.gov/~bottom/prescales`.

[37] The B Physics Analysis Kernel Groud (BPAK). *Recommended database flags for a good run selection.* `http://www-cdf.fnal.gov/internal/physics/bottom/bpak/`.

[38] K. Bloom and W.-M. Yao. *Outside-In Silicon Tracking at CDF*, 2002. CDF/DOC/TRACKING/CDFR/5991.

[39] K. Anikeev, G. Bauer, I. Furic, S. Gromoll, A. Korn, I. Kravchenko, M. Mulhearn, Ch. Paus, A. Rakitin, and J. Tseng. *Calibration of Energy Loss and Magnetic Field using J/ψ Events in Run II*, 2002. CDF/DOC/BOTTOM/CDFR/5958.

[40] J. Boudreau, C. Green, P. Murat, M. Paterno, and L. Sexton-Kennedy. *The CDF Geometry System for Run II Offline Software. A Guide for Subdetector Package Writers*, 2000. CDF/DOC/COMP_UPG/PUBLIC/5368.

[41] Mario Campanelli and Elena Gerchtein. *Calibration of the momentum scale for Kalman refitter using J/psi events.*, 2004. CDF/PHYS/BOTTOM//6905.

[42] Ch. Paus et al. see `http://cdfcodebrowser.fnal.gov/CdfCode/source/BottomMods`.

[43] C.Blocker, J.Boudreau, and Ch.Paus. *Common Tools for B Physics Analyses in Run II at CDF*, 2002. CDF/DOC/BOTTOM/CDFR/5735.

[44] J. Marriner. *SECONDARY VERTEX FIT WITH MASS AND POINTING CONSTRAINTS (CTVMFT)*, 1993. CDF/DOC/SEC_VTX/PUBLIC/1996.

[45] I. K. Furic. *Measurement of the Ratio of Branching Fractions* $Br(B_s^0 \to D_s\pi)/Br(B_d^0 \to D\pi)$, 2004. CDF/ANAL/BOTTOM/CDFR/7352.

[46] I. Furic, B. Iyutin, and Ch. Paus. *BottomMods – Stntuples for B physics*, 2005.

[47] Rene Brun and Fons Rademakers. *ROOT - An Object Oriented Data Analysis Framework*, 1997. Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also http://root.cern.ch/.

[48] J. Heuser. Private communication, 2006.

[49] CDF B_s Mixing Group. *Updated measurement of B^0 Oscillations and Calibration of Flavor Tagging in Fully Reconstructed Decays*, 2005. CDF/DOC/BOTTOM/CDFR/7722.

[50] A. Belloni et al. *An Update to the Measurements $Br(B_s \to D_s\pi)/Br(B^0 \to D\pi)$ and $Br(B^+ \to D\pi)/Br(B^0 \to D\pi)$*, 2004. CDF/DOC/BOTTOM/PUBLIC/7351.

[51] The CDF Collaboration. *Observation of Λ_b to Λ_c π and the measurement of f_{Λ_b}/f_d $(BR(\Lambda_b)/BR(B^0))$* , 2005. `hep-ex/0601003`.

[52] J. Piedra et al. *Unbinned Likelihood Fit for B^0 Mixing in Fully Reconstructed Decays*, 2004. CDF/DOC/BOTTOM/CDFR/7388.

[53] N. Leonardo et al. *Methods for Likelihood Computation in Lifetime and Mixing Analyzes*, 2004. CDF/ANAL/BOTTOM/PUBLIC/7385.

[54] A. Bolshov et al. *B mesons lifetime determination in fully reconstructed hadronic decays.*, 2005. CDF/DOC/BOTTOM/CDFR/7386.

[55] The CDF Collaboration. *Muon B Flavor Tagging - A Likelihood Approach*, 2005. CDF Public Notes 7043 and 7121, available at `http://www-cdf.fnal.gov/physics/new/bottom/bottom.html`.

[56] The CDF Collaboration. *A Neural Network based Jet Charge Tagger*, 2005. CDF Public Note 7482, available at `http://www-cdf.fnal.gov/physics/new/bottom/bottom.html`.

[57] CDF B_s Mixing group. *Same Side Kaon Tagging Studies on Fully Reconstructed Decays*, 2005. CDF Note 7979.

[58] CDF B_s Mixing group. *Validation of New Datasets for the B_s Oscillation Studies in $B_s \to D_s(3)\pi$ decays*, 2006. CDF/DOC/BOTTOM/CDFR/8172.

[59] CDF B_s Mixing Group. *Scale Factors for Proper Time Uncertainties with Event-by-Event Primary Vertex Reconstruction*, 2005. CDF Note 7818 and CDF Public Note 7944, available at `http://www-cdf.fnal.gov/physics/new/bottom/bottom.html`.

[60] A. Bolshov et al. *Measurement of the Ratio of Branching Fractions $Br(B_s \to D_s\pi)/Br(B^0 \to D\pi)$ with 355 pb^{-1}* , 2005. CDF Public Note 7862, available at `http://www-cdf.fnal.gov/physics/new/bottom/bottom.html`.

[61] C. Lecci et al. (CDF B_s Mixing Group). *Monte Carlo study of the lepton SVT sample*, 2004. CDF/PHYS/BOTTOM/CDFR/7318.

[62] D. J. Lange. The evtgen particle decay simulation package. Nucl. Instrum. Meth., A462:152–155, 2001.

[63] *The CDF Detector Simulation.* `http://www-cdf.fnal.gov/cdfsim/cdfsim_main.html`.

[64] B. Ashmanskas and A. Cerri. see `http://cdfcodebrowser.fnal.gov/CdfCode/source/svtsim`.

[65] E. Norrbin and T. Sjostrand. Production and hadronization of heavy quarks. Eur. Phys. J., C17:137–161, 2000.

[66] C. Dörr et al. *Signal Optimization for the decays $B_s \to D_s(3)\pi$ with Neural Networks*, 2006. CDF Note, to be pusblished.

[67] U. Kerzel. Neurobayes – a sophisticated neural network package. Talk at BPAK meeting Dec. 10, 2004.

[68] L. Ramler (phi-t GmbH). Private communication, 2006.

[69] Wolfram Research Inc. *The Wolfram Integrator.* `http://integrals.wolfram.com`.

[70] F. Bedeschi, M. Caldarelli, and D. Lucchesi. *Formulas to describe resolution effects in mixing analysis*, 1994. CDF/ANAL/BOTTOM/CDFR/2859.

Acknowledgements

I would like to thank Prof. Dr. Michael Feindt for accepting me as a PhD student in his group and for the guidance and the many instructive discussions about physics and statistics during my graduate work.
I would like to thank Prof. Dr. Günter Quast for his co–supervision of this thesis. Also I would like to thank Prof. Dr. Thomas Müller for making several visits to Fermilab possible and Mrs. Haas and Mr. Fuchs for facilitating the travel organization.
This work was made possible by the support of the Studienstiftung des deutschen Volkes and the Graduiertenkolleg Hochenergieteilchen– und Teilchenastrophysik. I thank them for the scholarships, workshops, language courses and summer academies that enriched my graduate studies.
I thank all members of the Karlsruhe B group for the great working atmosphere and the many fruitful discussions. In particular I want to thank Dr. Ulrich Kerzel, Dr. Michal Kreps and Dr. Kurt Rinnert for the invaluable help in all technical questions related to CDF software, C++, ROOT and SAM. I want to express my gratitude to Martin Heck, Joachim Heuser, Dr. Ulrich Kerzel, Dr. Michal Kreps, Dr. Thomas Kuhr, Jörg Lange, Philipp Mack, Michael Milnik, Svenja Richter and PD Dr. Wolfgang Wagner for carefully reading this manuscript and helping to improve it by providing many helpful comments and suggestions.
I would like to thank all people at the EKP that supported this work, in particular the always reliable and friendly Admin team that does a brilliant job in keeping all systems up and running.
Special thanks go to all people at Fermilab who helped me using the B Stntuple framework and directed me to the corresponding data samples. In particular I want to thank Dr. Alberto Belloni, Dr. Guillelmo Ceballos–Gomez and Dr. Stephanie Menzemer, without their help and guidance this work wouldn't have been possible.
I would like to thank all room–mates with whom I shared office 9–2 ("Diplomandenzimmer") over the years for the great time and the many interesting discussions about the important questions in life.
Last but not least I want to express my gratitude to my parents for their continuous help and support. I want to thank Claudia Robbers for her support and for being a constant source of inspiration.