

Study of Flavor Tagging with Semi-exclusive Reconstruction of
 $B \rightarrow D_{(s)}X$ Decays.

by

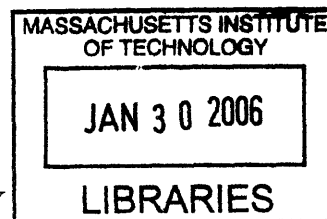
Elizabeth S. Ptacek

Submitted to the Department of Physics in partial fulfillment of the Requirements
for the Degree of

BACHELOR OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY



June, 2005

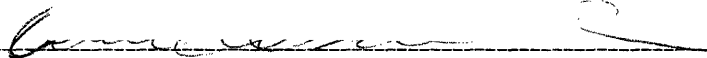
©2005 Elizabeth Ptacek

All Rights Reserved

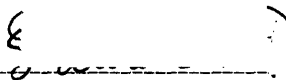
The author hereby grants to MIT permission to reproduce and to distribute publicly
paper and electronic copies of this thesis document in whole or in part.

Signature of Author -----

Department of Physics
(May 20, 2005)

Certified by -----

Professor Gabriella Sciolla
Thesis Supervisor, Department of Physics

Accepted by -----

Professor David E. Pritchard
Senior Thesis Coordinator, Department of Physics

ARCHIVES

Study of Flavor Tagging with Semi-exclusive Reconstruction of $B \rightarrow D_{(s)}X$ Decays.

by

Elizabeth S. Ptacek

Submitted to the Department of Physics on May 20, 2005 in partial fulfillment of
the Requirements for the Degree of Bachelor of Science in Physics

Abstract

The BaBar experiment measures CP violation in the B meson system by analyzing the decays of $B^0\bar{B}^0$ pairs in which one of the mesons decays into a CP eigenstate. By measuring the decay time distribution separately for B^0 and \bar{B}^0 decays the CP violating asymmetry and $\sin 2\beta$ can be measured. In order to measure the decay time distribution for the B^0 and \bar{B}^0 decays, it is necessary to identify the flavor, or tag, of the second B meson. The flavor of the B meson is identified by an algorithm which analyzes physical processes which carry flavor information from its decay. Analyzing new physical processes could potentially improve the efficiency of the tagging algorithm. I tested the effect of adding subtaggers which estimated the flavor of the tag B based on the flavor or the charge of its D^0 or D_s daughters. However due to the low yields and high backgrounds of reconstructed D^0 and D_s candidates, these subtaggers were not effective at making a net contribution to the tagging process.

Thesis Supervisor: Gabriella Sciolla

Title: Assistant Professor

Contents

1	Introduction	8
2	Theory	9
2.1	Symmetries of Nature	9
2.2	Parity Violation	9
2.3	CP Violation in the Neutral K Meson System	10
2.4	CP Violation in the Quark Sector	11
2.5	CP Violation in the B-meson System	12
2.6	CPT Theorem	14
2.7	The Excess of Matter Created in the Big Bang	14
3	The Accelerators	16
4	The BaBar Detector	17
4.1	The SVT	17
4.2	The DCH	18
4.3	The DIRC	18
4.4	The EMC	19
4.5	The IFR	19
5	Data Processing	20
5.1	The Trigger System	20
5.2	Reconstruction	20
5.2.1	Pions	21
5.2.2	Electrons	21
5.2.3	Muons	21
5.2.4	Kaons	22
5.2.5	Photons, π^0 s, and η s	22
5.2.6	Neutrinos	22
5.3	Candidate Reconstruction	22
6	Tagging	24
6.1	The D^0 and D_s Subtaggers	24
6.2	Input Variables for the D^0 Subtagger	25
6.3	Input Variables for the D_s^\pm Subtagger	25

6.4	The Effective Tagging Efficiency (Q)	32
6.5	The Results	34
6.5.1	Correlation Between Inputs and Output for the D^0 and D_s Subtaggers	34
6.5.2	D^0 Subtagger Performance	39
6.5.3	D_s Subtagger Performance	42
6.5.4	BTagger Performance	42
7	Conclusion	45
8	Acknowledgments	45
A	Neural Networks	48
A.1	Separating Correlated Inputs	48
A.2	Training	51

List of Figures

1	Feynman diagrams of neutral kaon mixing.	10
2	The Unitarity Triangle[11].	12
3	Left: Distribution of $ktag$ for signal from right sign $B^0 \rightarrow \bar{D}^0 X$ in red; for background from wrong sign $B^0 \rightarrow D^0 X$ in blue and for combinatorial background in black. Right: purity $(\frac{signal}{background+signal})$ versus $ktag$	26
4	Left: Distribution of $pstar$ for signal from right sign $B^0 \rightarrow \bar{D}^0 X$ in red; for background from wrong sign $B^0 \rightarrow D^0 X$ in blue and for combinatorial background in black. Right: purity $(\frac{signal}{background+signal})$ versus $pstar$	27
5	Left: Distribution of the invariant $mass$ for signal from right sign $B^0 \rightarrow \bar{D}^0 X$ in red; for background from wrong sign $B^0 \rightarrow D^0 X$ in blue and for combinatorial background in black. Right: purity $(\frac{signal}{background+signal})$ versus $mass$	27
6	Left: Distribution of $vtxprob$ for signal from right sign $B^0 \rightarrow \bar{D}^0 X$ in red; for background from wrong sign $B^0 \rightarrow D^0 X$ in blue and for combinatorial background in black. Right: purity $(\frac{signal}{background+signal})$ versus $vtxprob$	28
7	Left: Distribution of $costheta$ for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity $(\frac{signal}{background+signal})$ versus $costheta$	28

8	Left: Distribution of $vtxprob$ for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity $(\frac{signal}{background+signal})$ versus $vtxprob$	29
9	Left: Distribution of $pstar$ for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity $(\frac{signal}{background+signal})$ versus $pstar$	29
10	Left: Distribution of the invariant $mass$ for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity $(\frac{signal}{background+signal})$ versus $mass$	30
11	Left: Distribution of $d1massdif$ in $D_s^+ \rightarrow \phi\pi^+$ decays for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity versus $d1massdif$	30
12	Left: Distribution of $d1massdif$ in $D_s^+ \rightarrow K^+\bar{K}^0(s)$ decays for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity versus $d1massdif$	31
13	Left: Distribution of $d1massdif$ in $D_s^+ \rightarrow \bar{K}^*(892)^0 K^+$ decays for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity versus $d1massdif$	31
14	For the D^0 subtagger the distribution of $ktag$ values for candidates with output values within a given interval divided by the $ktag$ values with no cut on the output; blue: output > 0.7, cyan: 0.7 > output > 0.4, green: 0.4 > output. For reference the purity versus $ktag$ is graphed in black to the right.	34
15	For the D^0 subtagger the distribution of input variable values for candidates with output values within a given interval divided by the input variable values with no cut on the output; blue: output > 0.7, cyan: 0.7 > output > 0.4, green: 0.4 > output. For reference the purity versus each input variable is graphed in black to the right.	35
16	For the D_s subtagger the distribution of input variable values for candidates with output values within a given interval divided by the input variable values with no cut on the output; blue: output > 0.4, cyan: 0.4 > output > 0.2, green: 0.2 > output. For reference the purity versus each input variable is graphed in black to the right.	36

- 17 For the D_s subtagger the distribution of $vtxprob$ values for candidates with output values within a given interval divided by the $vtxprob$ values with no cut on the output; blue: output > 0.4, cyan: $0.4 > \text{output} > 0.2$, green: $0.2 > \text{output}$. For reference the purity versus $vtxprob$ is graphed in black to the right. 37
- 18 For the D_s subtagger the distribution of $d1massdif$ values for candidates from $D_s^+ \rightarrow \phi\pi^+$ decays with output values within a given interval divided by the $d1massdif$ values with no cut on the output; blue: output > 0.4, cyan: $0.4 > \text{output} > 0.2$, green: $0.2 > \text{output}$. For reference the purity versus $d1massdif$ is graphed in black to the right. 37
- 19 For the D_s subtagger the distribution of $d1massdif$ values for candidates from $D_s^+ \rightarrow K^+\bar{K}^0(s)$ decays with output values within a given interval divided by the $d1massdif$ values with no cut on the output; blue: output > 0.4, cyan: $0.4 > \text{output} > 0.2$, green: $0.2 > \text{output}$. For reference the purity versus $d1massdif$ is graphed in black to the right. 38
- 20 For the D_s subtagger the distribution of $d1massdif$ values for candidates from $D_s^+ \rightarrow \bar{K}^*(892)^0 K^+$ decays with output values within a given interval divided by the $d1massdif$ values with no cut on the output; blue: output > 0.4, cyan: $0.4 > \text{output} > 0.2$, green: $0.2 > \text{output}$. For reference the purity versus $d1massdif$ is graphed in black to the right. 38
- 21 Distribution of the output of the neural network for the D^0 subtagger for candidates from B^0 decays in blue and \bar{B}^0 decays in red. 39
- 22 Left: Distribution of the output from the D^0 subtagger for B^0 decays with $pstar > 1.7$ in red and $pstar > 0$ in black. Right: Purity vs. $pstar$ with same color coding. 39
- 23 Distribution of the output from D^0 subtagger for candidates from B^0 decays. Signal is graphed in blue and background in red. 40
- 24 Distribution of the output of the neural network for the D_s subtagger for candidates from B^0 decays in blue and \bar{B}^0 decays in red. 41
- 25 Left: Distribution of the output from the D_s subtagger for B^0 decays with $pstar > 1.3$ in red and $pstar > 0$ in black. Right: Purity vs. $pstar$ with same color coding. 41
- 26 Distribution of the output from D_s subtagger for candidates from B^0 decays. Signal is graphed in blue and background in red. 42

27	Distribution of the output of BTagger without the D^0 or D_s subtaggers included for $B^0 \rightarrow \bar{D}^0 X$ decays where the \bar{D}^0 was correctly reconstructed. .	43
28	Distribution of the output of BTagger without the D^0 or D_s subtaggers included for $B^0 \rightarrow D_s^+ X$ decays where the D_s^+ was correctly reconstructed.	43
29	Graph of linearly separable sets A and B	48
30	Hyperplane separating sets A and B	49
31	Graph of two sets A and B which are not linearly separable.	49
32	Node configuration which separates sets graphed in figure 31.	50
33	(Left) Two arbitrary sets which are separable by a neural net with one hidden layer.	50
34	(Right) A neural net with one hidden layer of m nodes which can be used to separate the two sets graphed to the left.	50
35	A neural network with $n+1$ inputs, $m+1$ nodes and $m+1$ outputs - one for each node.	51
36	A neural network with $n+1$ inputs, a single layer of $m+1$ hidden nodes, and one output.	52

List of Tables

1	Table of decay modes reconstructed for D^0 and D_s^\pm candidates	23
---	---	----

1 Introduction

The BaBar experiment located at the Stanford Linear Accelerator Center measures the CP violating asymmetry in the neutral B meson system by analyzing events in which a $B^0\bar{B}^0$ pair is produced and one of the B mesons (B_{CP}) decays to a CP eigenstate. The other B meson (B_{tag}) is identified as either a B^0 or a \bar{B}^0 in a process called tagging in which a number of physical processes which carry flavor information about the tag B are analyzed. The sensitivity to CP violation depends on the effectiveness of the tagging algorithm. My research explored improving the tagging algorithm by exploiting the flavor and charge of the D^0 and D_s daughters of the tag B as an additional basis for the tag.

In this paper I discuss the theory of CP violation and its measurement at BaBar with particular emphasis on the tagging algorithm. Section 2 introduces CP violation discussing how it was discovered, how it arises in the quark sector of the Standard Model, how the CP violating asymmetry can be computed from measurements made by BaBar, and the theoretical consequences of CP violation. Sections 3 and 4 cover the accelerators and the BaBar detector. Section 5 summarizes reconstruction of particle species. Section 6 covers the tagging algorithm, BTagger, and the two subtaggers which I added, the inputs, the training of the neural networks, and the correlation between the inputs and the output of these subtaggers. Finally the results of adding the two new subtaggers to BTagger is discussed.

2 Theory

2.1 Symmetries of Nature

In the Standard Model of particle physics the C, P and T symmetries are of fundamental importance. Charge conjugation, C, reverses the sign of all internal quantum number such as charge, beauty, strangeness, etc. It thus converts a particle into its antiparticle. Parity, P, reverses all spatial coordinates sending $x \rightarrow -x$, $y \rightarrow -y$, and $z \rightarrow -z$. This is essentially equivalent to converting a particle into its mirror image. The parity operator reverses the right- or left-handedness of a particle. A particle is by definition right-handed when its spin direction, as determined by the right-hand rule, is along the direction of its motion and left-handed when its spin direction is opposite its direction of motion. Time reversal, T, reverses the direction of time for a process.

2.2 Parity Violation

It was originally believed that the Standard Model should be invariant under all three symmetries C, P, and T. In fact, in all experiments done to date strong, electromagnetic, and gravitational interactions are invariant under the C, P, and T operators. However weak interactions are not invariant under C and P. The possibility of parity violation was postulated in 1956 by Lee and Yang[16] and observed the same year by C.S. Wu[20], who showed that the electrons emitted in the beta decay of Cobalt 60 nuclei are preferentially emitted in the direction of nuclear spin[10]. This result is due to the fact that beta decay occurs via weak interactions and in weak interactions the C and P symmetries are maximally violated. A typical weak decay is $\pi^+ \rightarrow \mu_L^+ \nu_L$. Applying C to this decay results in $\pi^- \rightarrow \mu_L^- \bar{\nu}_L$. However such a decay is never observed in nature. There are no left-handed antineutrinos. Also if P is applied to the decay $\pi^+ \rightarrow \mu_L^+ \nu_L$, then the result is $\pi^+ \rightarrow \mu_R^+ \nu_R$ which also is never observed since there are no right-handed neutrinos in nature. Applying both C and P results in the decay which is actually experimentally observed: $\pi^- \rightarrow \mu_R^- \bar{\nu}_R$. In fact, weak interactions are almost symmetric under the combined operator, CP. However they are not perfectly symmetric.

2.3 CP Violation in the Neutral K Meson System

CP violation was first observed in 1964 by Cronin and Fitch[8] in the decays of K_L^0 mesons into two pion states. The neutral kaons have flavor eigenstates $K^0 = d\bar{s}$ and $\bar{K}^0 = s\bar{d}$ which can turn into each other (mix) as illustrated by the diagrams in figure 1.

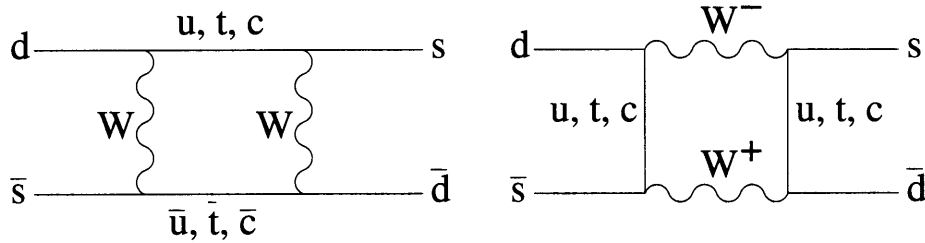


Figure 1: Feynman diagrams of neutral kaon mixing.

Their CP eigenstates can be written as:

$$|K_1\rangle = \frac{1}{\sqrt{2}}(|K^0\rangle - |\bar{K}^0\rangle) \text{ and } |K_2\rangle = \frac{1}{\sqrt{2}}(|K^0\rangle + |\bar{K}^0\rangle)$$

$|K_1\rangle$ undergoes the two pion decays $|K_1\rangle \rightarrow \pi^+\pi^-, \pi^0\pi^0$ and has a CP eigenvalue of +1; $|K_2\rangle$ undergoes the three pion decays $|K_2\rangle \rightarrow \pi^+\pi^-\pi^0, \pi^0\pi^0\pi^0$ and has a CP eigenvalue of -1. If CP were a perfect symmetry, then these CP eigenstates of the neutral K meson, K_1 and K_2 , would be the same as the mass eigenstates of the neutral K meson, K_S^0 and K_L^0 . This hypothesis was proved false by the Cronin and Fitch experiment. In the Cronin-Fitch experiment[8] a beam of neutral K mesons was created and then the K_L^0 's were isolated by sending the beam over a distance of 57 feet. Since the K_S^0 has a lifetime which is about 584 times shorter than that of the K_L^0 , the K_S^0 's decayed out of the beam. The experiment proved that about 1 in every 500 K_L^0 's in the purified beam decayed into two pions, a CP eigenstates of eigenvalue +1, demonstrating that the K_L^0 contains a small K_1 component. Further evidence of CP violation, this time in the B-meson system, was discovered in 2001 by the BaBar[1] and Belle[5] experiments. I will discuss CP violation in the B-meson system in more detail in section 2.5. The following section is devoted to discussing how CP violation arises in the Standard Model.

2.4 CP Violation in the Quark Sector

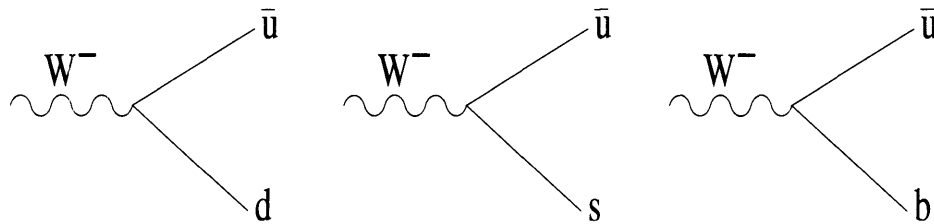
In the Standard Model, CP violation arises as a result of the existence of a complex phase in the CKM matrix which describes the flavor changing weak interactions of quarks. The W and Z bosons, the mediators of the weak force, couple with each other, leptons, neutrinos, and quarks. The quark states with which the weak force couples however are not the physical mass eigenstates but instead a superimposition of these states. The weak force couples the following pairs:

$$\begin{pmatrix} u \\ d' \end{pmatrix} \begin{pmatrix} c \\ s' \end{pmatrix} \begin{pmatrix} t \\ b' \end{pmatrix} \quad (1)$$

where d' , s' , and b' are linear combinations of d , s , and b . The weak interactions of quarks are described by the following matrix known as the CKM (Cabibbo-Kobayashi-Maskawa) matrix[12]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2)$$

The magnitude of each matrix element gives the strength of the transitions between the corresponding physical quark states. For instance V_{ud} , V_{us} , and V_{ub} measure the coupling for the following vertexes respectively:



The experimentally determined magnitudes of the elements of the CKM matrix are[10]:

$$\begin{pmatrix} 0.9705 - 0.9770 & 0.21 - 0.24 & 0 - 0.014 \\ 0.21 - 0.24 & 0.971 - 0.973 & 0.036 - 0.070 \\ 0.0 - 0.024 & 0.036 - 0.069 & 0.997 - 0.999 \end{pmatrix} \quad (3)$$

It can be shown using unitary constraints and the arbitrary nature of the relative quark phases that the nine elements of the CKM matrix can be rewritten in terms

of four real parameters. Wolfenstein proposed the following parameterization for the CKM matrix[19]:

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

The CKM matrix is unitary. This leads to six relations involving the nine matrix elements. One of these relations $V_{ub}V_{ud}^* + V_{cb}V_{cd}^* + V_{td}V_{tb}^* = 0$ is particularly useful. It implies that the vectors $V_{ub}V_{ud}^*$, $V_{cb}V_{cd}^*$, and $V_{td}V_{tb}^*$ form a triangle in the complex plane. This triangle, which is normalized by dividing the side lengths by $V_{cd}V_{cb}^*$, is known as the Unitarity Triangle.

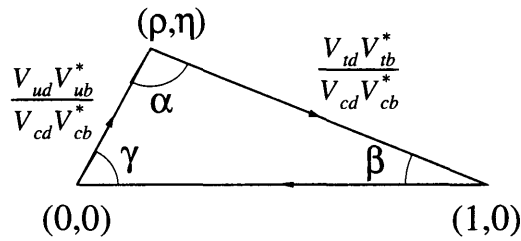


Figure 2: The Unitarity Triangle[11].

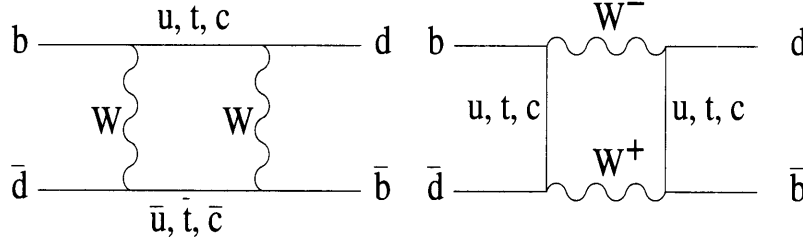
Measuring the height of this triangle, the CKM phase (η), is particularly important since in the Standard Model a nonzero η implies the existence of CP violation. All sides and angles of the Unitarity Triangle can be measured in the decay of B mesons. In particular, $\sin 2\beta$ can be measured from the decays of neutral B mesons with a precision better than 4%. A non-null value for $\sin 2\beta$ implies a non-null value for the CKM phase.

2.5 CP Violation in the B-meson System

In the BaBar experiment $B^0\bar{B}^0$ pairs are produced from an $\Upsilon(4S)$ resonance. A B^0 meson is composed of an anti-bottom quark and a down quark. Its mass eigenstates $|B_L\rangle$ and $|B_H\rangle$ can be written as linear combinations of its flavor eigenstates[11]:

$$\begin{aligned} |B_L\rangle &= p |B^0\rangle + q |\bar{B}^0\rangle \\ |B_H\rangle &= p |B^0\rangle - q |\bar{B}^0\rangle \end{aligned} \quad (5)$$

where p and q are complex coefficients. $|q|^2 + |p|^2 = 1$. The B^0 meson undergoes regular transitions between flavor eigenstates as represented by the Feynman diagrams:



The equations for the time evolution of a pure B^0 or \bar{B}^0 state at $t = 0$ are[11]:

$$\begin{aligned} |B^0(t)\rangle &= e^{-im_B t} e^{-\frac{t}{2\tau_{B^0}}} \left(\cos \frac{\Delta m t}{2} |B^0\rangle - i \frac{q}{p} \sin \frac{\Delta m t}{2} |\bar{B}^0\rangle \right) \\ |\bar{B}^0(t)\rangle &= e^{-im_B t} e^{-\frac{t}{2\tau_{B^0}}} \left(\cos \frac{\Delta m t}{2} |\bar{B}^0\rangle - i \frac{p}{q} \sin \frac{\Delta m t}{2} |B^0\rangle \right) \end{aligned} \quad (6)$$

where $m_B = \frac{(m_H + m_L)}{2}$, m_H and m_L are the masses of the mass eigenstates, and τ_{B^0} is the lifetime of the B^0 meson¹. The $B^0\bar{B}^0$ state evolves coherently; by the laws of quantum mechanics, when one of these particles is a B^0 the other is always its antiparticle, a \bar{B}^0 , until one of the two mesons decays. Because of this, events in which one of the B^0 mesons decays to a CP eigenstate and the other, the tag B, decays to a state of identifiable flavor can be used to measure the CP-violating asymmetry in the B meson system. For such events, interference between the mixing induced amplitude ($B^0 \rightarrow \bar{B}^0 \rightarrow f_{CP}$) and the decay amplitude ($B^0 \rightarrow f_{CP}$) to a CP eigenstates f_{CP} leads to a time dependent CP asymmetry[1]. This asymmetry is experimentally measurable and a function of the angles in the unitary triangle. The distribution $f_{B^0}(f_{\bar{B}^0})$ of the decay rate when the tagging meson is a $B^0(\bar{B}^0)$ is given by[1]:

$$f_{B^0}(\Delta t) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{2\tau_{B^0}(1 + |\lambda_f|^2)} \left(\frac{1 + |\lambda_f|^2}{2} + Im\lambda_f \sin(\Delta m_{B^0} \Delta t) - \frac{1 - |\lambda_f|^2}{2} \cos(\Delta m_{B^0} \Delta t) \right) \quad (7)$$

$$f_{\bar{B}^0}(\Delta t) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{2\tau_{B^0}(1 + |\lambda_f|^2)} \left(\frac{1 + |\lambda_f|^2}{2} - Im\lambda_f \sin(\Delta m_{B^0} \Delta t) + \frac{1 - |\lambda_f|^2}{2} \cos(\Delta m_{B^0} \Delta t) \right) \quad (8)$$

¹It is assumed that the lifetime difference between the two flavor eigenstates is negligible.

The flavor of the tag B meson is identified in a process called tagging which will be discussed in detail in subsequent sections. Δm_{B^0} is the mass difference between the flavor eigenstates of the neutral B meson. Δt is the time interval between the decays of the two mesons as measured in the rest frame of $\Upsilon(4S)$. It is extracted from the distance between the two decay vertexes, $\Delta z = \beta\gamma\Delta t$ with $\beta\gamma \approx 0.56c$ at BaBar. On average Δt is around 1.5×10^{-12} sec corresponding to a Δz of around $250\mu m$. λ_f is a complex parameter defined as $\lambda_f = \eta_f \frac{q}{p} \frac{\bar{A}_f}{A_f}$, where A_f and \bar{A}_f are the decay amplitudes of the B^0 and \bar{B}^0 mesons into the CP eigenstate f_{CP} :

$$A_f = \langle f_{CP} | H | B^0 \rangle \text{ and } \bar{A}_f = \langle f_{CP} | H | \bar{B}^0 \rangle.$$

When the CP eigenstate is $J/\psi K_S^0$, $\psi(2S)K_S^0$ or $J/\psi K_L^0$, the Standard Model predicts that $\lambda_f = \eta_f e^{-2i\beta}$. The CP eigenvalue η_f is -1, -1 or +1 for the CP eigenstates $J/\psi K_S^0$, $\psi(2S)K_S^0$ or $J/\psi K_L^0$ respectively. So $|\lambda_f| = 1$ and $\sin 2\beta$ follows by a simple calculation from the measurement of the decay rate distributions and Δt :

$$A_{CP}(\Delta t) = \frac{f_{B^0}(\Delta t) - f_{\bar{B}^0}(\Delta t)}{f_{B^0}(\Delta t) + f_{\bar{B}^0}(\Delta t)} = -\eta_f \sin 2\beta \sin(\Delta m_{B^0} \Delta t) \quad (9)$$

The best measurement of $\sin 2\beta$ made by BaBar to date is $0.722 \pm 0.040 \pm 0.023$. The value measured by Belle is $0.728 \pm 0.056 \pm 0.023$ which is consistent with BaBar's estimate. The world average for $\sin 2\beta$ is 0.725 ± 0.037 providing conclusive evidence for CP violation in the B meson system[11].

2.6 CPT Theorem

CP violation has a number of consequences which are of fundamental importance. In current theory, CPT is a symmetry of nature. In fact it can be shown that it is impossible to construct a quantum field theory in which CPT is violated. If CPT is in fact not violated, then T must be violated to compensate for CP violation, or, in other words, fundamental processes of nature must not be strictly invariant under time reversal.

2.7 The Excess of Matter Created in the Big Bang

CP violation is necessary to explain the existence of the universe. If there were a perfect symmetry between the worlds of matter and antimatter, as would be the case if either C or CP were a perfect symmetry of nature, then equal amounts of matter

and antimatter should have been produced in the Big Bang. These equal quantities of matter and antimatter would have annihilated each other completely. Measurements of cosmic background radiation and evaluation of baryonic matter in the universe indicate that for about every billion matter and antimatter particles produced there was one extra matter particle. As first postulated by Sakharov in 1967, CP violation could account for this slight excess of matter over antimatter[18].

3 The Accelerators

In the BaBar experiment located at SLAC (the Stanford Linear Accelerator Center), electron and positron beams are first accelerated by a linear accelerator and then sent into the PEP II storage rings where they undergo collisions in the center of the BaBar detector. The beam tunnel for the linear accelerator is two miles long and is buried 25 feet below the earth's surface. Above the beam tunnel is the klystron gallery. The klystrons generate microwave frequency electromagnetic radiation which is channeled down to the linac by wave guides. The linac has a copper structure which is divided into cavities. The microwave radiation causes electromagnetic waves to travel down these cavities. The electrons and positrons which follow each other in bunches down the linac are accelerated by these waves since their velocity is synchronized with the oscillations of the electromagnetic fields. Also the electric field of the wave reverses direction every three cavities so the spacing between the electron and positron bunches must be a multiple of three cavities in length. The electrons reach a final energy of 9.0 GeV while the positrons reach a final energy of 3.1 GeV. At the end of the accelerator a dipole magnet is used to separate the electrons and the positrons. The electron and positron beams are then directed into the rings of the PEP II facility.

PEP II has two independent storage rings one on top of the other. The electron and positron beams cycle in opposite directions each in its own ring. To compensate for energy losses due to synchrotron radiation the electrons and positrons are accelerated as they cycle in the PEP II rings, and thus the beams are maintained at their respective energies of 9.0 and 3.1 GeV. In the center of the BaBar detector the electron and positron beams are sent through each other. The center of mass energy of electrons and positrons which collide from these two beams is 10.58 GeV just above the threshold for producing a $B^0\bar{B}^0$ pair, the mass of the $\Upsilon(4S)$ resonance.

4 The BaBar Detector

The BaBar detector is designed to record the decay products of the B mesons. The detector is asymmetric; it is designed to be more sensitive to decay products traveling in the forward direction, the direction of the higher energy electron beam. This detector is composed of a number of subdetectors. Innermost around the beam pipe is the cylindrical silicon vertex tracker (SVT) which extends from an inner radius of a few centimeters out to about 20 centimeters. Surrounding the SVT is the drift chamber (DCH) with an inner radius of 24 cm and an outer radius of 81 cm. Around the DCH is the Detector of Internally Reflected Cherenkov radiation (DIRC). Outside the DIRC is the electromagnetic calorimeter which is composed of a barrel region with an inner radius of 90 cm and an outer radius of 138 cm and a forward endcap. A superconducting solenoid surrounds these detectors and generates an interior magnetic field of 1.5 tesla. Outermost is the steel shell of the instrumented flux return (IFR) which acts as a flux return for this magnet as well as a detector for muon and K_L identification.

4.1 The SVT

The SVT is composed of five layers of double sided silicon strip sensors[2]. The sensors are organized in modules which are arranged in five cylindrical layers ranging in radii from 3 to 11 cm. The ends of the two outermost layers are arched inwards towards the beam pipe. When a charged particle passes through one of the strips, it ionizes the silicon creating electron-hole pairs. An electric field is applied across the silicon so that the electrons and holes migrate creating a current that is sensed as a hit on the detector. The SVT is capable of recording the passage of any charged particle. Its principal purpose is to locate the decay vertexes of very short-lived particles. It has a high resolution, $\sim 15\mu m$, which is essential since it allows for a precise measurement of the time-dependent CP asymmetry. The SVT also allows for the tracks of particles which never reach the DCH to be reconstructed, and measures the ionization $\frac{dE}{dx}$. The measurement of $\frac{dE}{dx}$ provides PID information which is used to differentiate between kaons, pions and protons.

4.2 The DCH

Surrounding the SVT is the drift chamber. The drift chamber is formed from 7,104 hexagonal cells arranged in 40 layers. The cells have extents of $1.2 \times 1.7 \text{ cm}^2$. Each cell consists of one sense wire surrounded by six field wires. In 16 of the layers the wires are strung along the z axis, while in 24 of the layers they are strung at small angles to the z axis in order to provide longitudinal resolution. The DCH is filled with a mixture of 80% Helium and 20% Isobutane gas. When a charged particle passes through the DCH it ionizes the gas. Since the sense wires are kept at a higher potential than the field wires ($\sim 2000 \text{ V}$ vs. 350 or 0 V), the positive ions drift towards the field wires and the electrons towards the sense wires. The readout electronics mounted at the rear of the detector measure the time interval between the ionization and the arrival of the electrons at the anode. From this time interval the position of primary ionization in the cell is obtained. The time intervals measured from multiple cells with different wire orientations can be combined to pinpoint the track's location. The primary function of the DCH is to measure the momenta and angles of charged particles[2]. The momentum is obtained from the track's curvature. In addition the ionization loss $\frac{dE}{dx}$ for a track is calculated from the net charge deposited in each cell. This quantity is used for particle identification for lower momentum tracks. In particular the $\frac{dE}{dx}$ measurements made by the DCH are used to separate kaons and pions with momenta less than $700 \text{ MeV}/c$.

4.3 The DIRC

The DIRC is located between the DCH and the EMC. It is roughly cylindrical in shape. It is composed of 144 bars of fused silica $1.7 \times 3.5 \times 490 \text{ cm}$. They lie lengthwise parallel to the beam pipe. In the plane perpendicular to the beam pipe the bars form a twelve sided polygon with side length of about 43 cm . Each side of the polygon is actually a bar box in which twelve bars are mounted side by side. Charged particles traveling through the silica at speeds greater than the speed of light in the silica emit Cherenkov radiation. This radiation is transmitted via internal reflection down the quartz bar to the rear end of the detector. There it is directed through the water filled standoff box to an array of PMTs (photo multiplier tubes). The internal reflection in the quartz preserves the emission angle of the Cherenkov light; therefore it is possible to reconstruct from the location and timing of the hits on the PMTs an image of the Cherenkov light cone. From the cone opening angle and the

measurement of the track's angle and momentum made in the DCH the mass of the particle can be calculated. Principally, the DIRC distinguishes between kaons and pions for energies from 500 MeV/c up to 4.5 GeV/c since the higher momentum kaons are not well identified by the $\frac{dE}{dx}$ measurement made in the DCH.

4.4 The EMC

Surrounding the DIRC is the electromagnetic calorimeter. It is made from 6580 Thallium doped Cesium iodide crystals. Of these 5760 are arranged in rings around the beam pipe which are formed into a cylinder. The rest are fashioned into rings which form a conically shaped endcap at the front of the detector. The crystals are oriented so that they point towards the interaction point. When a particle or photon travels through the crystals it creates an electromagnetic shower by scattering electrons. These electrons are knocked into an excited state. They then fall back down to the valance band emitting light in the process. The doping facilitates this transition and reduces the frequency of the emitted light by creating metastable states between the excited state and the ground state. The scintillation light emitted is detected by photodiodes mounted at the ends of the crystals, the outer surface of the calorimeter. The shape and magnitude of the shower created by a particle in the EMC is used for PID. The EMC's main purpose is to reconstruct photons and identify electrons. It also complements the IFR in the detection of muons and neutral hadrons.

4.5 The IFR

The IFR is the outermost shell of the detector. It is formed from a hexagonal barrel region and two endcaps of segmented steel. The steel shell of the IFR is composed of 18 layers ranging from 2 cm in thickness for the innermost to 10 cm in thickness for the outermost. Between these layers there are layers of single gap resistive plate chambers (RPCs) inserted. The RPCs consist of two bakelite, phenolic polymer, electrodes separated by a gap of a few millimeters. The gap is filled with a mixture of 57% argon, 39% freon, and 4% isobutane gas. High voltage is applied across the bakelite electrodes creating an electric field in the gap. When a particle flies through the detector it ionizes the gas changing the electric field and causing a detectable change in the capacitance of the RPC. The IFR serves a double purpose. First it acts as a flux return for the solenoid, and secondly it serves as a muon and neutral hadron detector.

5 Data Processing

5.1 The Trigger System

The trigger system is the first in a series of systems which process data from the detector. Its purpose is to filter the data so that detector readout is only recorded for events of physics interest such as $B^0\bar{B}^0$ production. The trigger system has two components: a hardware-based Level 1 trigger system and a software-based Level 3 trigger system. Output from the detector goes first into the Level 1 trigger system. Level 1 reconstructs tracks in the DCH and clusters in the EMC for each event. Based upon the number and quality of charged tracks and neutral centers, Level 1 decides whether or not to allow the data to pass on to the next level, the Level 3 trigger system. Level 3 further refines the reconstruction done by Level 1 reconstructing the entire event. It classifies the events and filters them based upon their classification to reduce the number of events of limited physics interest, such as Bahaba scattering, logged to disk.

5.2 Reconstruction

Once an event passes the trigger system it is given to the reconstruction software, a software package called Elf. The reconstruction software can also be run on a Monte Carlo generated simulation of raw detector output. The package that does this, including the simulation, is called Bear. The same reconstruction algorithm is run in either case. The first step in this algorithm is the reconstruction of tracks. The reconstruction software fits helixal paths to the DCH tracks which were reconstructed by the Level 3 trigger system. The tracks which can be successfully fitted are refined and any additional hits which lie on the helixal path are attributed to them. Next the software looks for additional tracks not detected by the Level 3 trigger system made solely of hits not previously assigned. Then all of the tracks are refitted using an improved algorithm which takes into account variations in the detector's material composition and magnetic field. The tracks are extrapolated into the SVT and again refit including the hits in the SVT. Tracks which do not exit the SVT are also reconstructed. Also tracks from particles which have scattered off the support tube between the SVT and the DCH are reconstructed taking such scattering into account. The final step in track fitting is performed after the particles have been identified. The procedures used to identify each type of elementary particle are specific to that

particle.

5.2.1 Pions

About 90% of the tracks generated in BaBar are pions, so if a track can not be positively identified as a kaon, proton, electron or muon, it is by default assigned to a pion list. Tracks which do not exit the SVT (transverse momentum $p_t < 120 \frac{MeV}{c}$) are also assumed to be pions. These particles are often slow pions coming from the decay $D^{*+} \rightarrow D^0\pi^+$. If the pion has a momentum p_t greater than $50 \frac{MeV}{c}$, it is reconstructed. Reconstruction efficiency for slow pions is greater than 80% for pions with momenta greater than $70 \frac{MeV}{c}$ [4]. The curvature of the pion's track in the magnetic field is used to measure its momentum.

5.2.2 Electrons

Electrons are identified based upon information from the EMC, DCH, and DIRC. The principle quantity used to identify electrons is the ratio $\frac{E}{p}$ of the energy deposited in the EMC to the momentum measured in the DCH and SVT. The energy loss ($\frac{dE}{dx}$) measured in the DCH is also used, as are the Cherenkov angle measured in the DIRC and the lateral shower development measured in the EMC. Depending upon the selector, the efficiency of the reconstruction program for electrons is either $\sim 95\%$ with a pion misidentification probability of $\sim 1.2\%$, or $\sim 88\%$ with a pion misidentification probability of $\sim 0.3\%$ [4].

5.2.3 Muons

The IFR is of primary importance in identifying muons. Tracks which pass out of the DCH are extrapolated and matched to hits in the IFR. Hits in several contiguous layers of the IFR are required for a particle to be identified as a muon. The energy deposited in the EMC and the degree of interaction of the muon candidate with other detector components is used to estimate the probability that it is a muon. Muons tend to interact less with detector components and deposit less energy in the EMC comparatively than pions. The efficiency in identifying muons with momenta above $1 \frac{GeV}{c}$ is around 77% with a pion fake rate of 2.5%[4].

5.2.4 Kaons

Charged kaons are identified using data from the SVT, DCH, and DIRC. The angle and intensity of the Cherenkov radiation measured in the DIRC and the energy loss ($\frac{dE}{dx}$) measured in the SVT and DCH are used to estimate the probability that a track is from a charged kaon. Charged kaons can be identified with an efficiency of around 90% and an average pion misidentification rate of around 2.5%[4].

K_L^0 s are identified by clusters in the EMC and hits in the IFR with no corresponding charged tracks in the drift chamber.

K_S^0 s undergo the decays $K_S^0 \rightarrow \pi^+\pi^-$ and $K_S^0 \rightarrow \pi^0\pi^0$. Therefore they are reconstructed either from two tracks of opposite sign with a common vertex, corresponding to the decays of the form $K_S^0 \rightarrow \pi^+\pi^-$, or from two π^0 candidates which satisfy the constraint that they could have been emitted by a particle of the kaon's mass, corresponding to decays of the form $K_S^0 \rightarrow \pi^0\pi^0$.

5.2.5 Photons, π^0 s, and η s

Photons are identified by clusters of hits in the EMC with an electromagnetic shower development and no corresponding tracks. π^0 s and η s are reconstructed from pairs of photons. For the purposes of this reconstruction the photons are assumed to have originated in the vicinity of the interaction point.

5.2.6 Neutrinos

The total momentum of all the particles generated in the decays of the B mesons should be zero in the $\Upsilon(4S)$ frame. If there is a net momentum, this suggests the production of neutrinos.

5.3 Candidate Reconstruction

Once the final state particle candidates have been identified, their tracks are again refit this time using their assigned masses. Then the reconstruction software begins reconstructing their mothers, grand mothers, etc. The software looks for sets of tracks which originate in a single vertex. For neutral particles it is assumed that they originated in the vicinity of the interaction point. The software then looks at the PID and kinematic information for the particles that came from this vertex in order to reconstruct the invariant mass of their mother. The daughter particles must be

consistent with the decay mode of the mother candidate. Also the computed mass of the reconstructed mother must be within a certain mass window of the mass of the mother candidate. Reconstructed particles, however, may not be real particles. Due to the large number of particles produced in any given event, a random collection of particles can appear to originate from the same vertex and be reconstructed by the software into a particle candidate that never actually existed. This combinatorial background is more common for decay modes with high multiplicity and a large number of pions or neutrals in the final state; it can be reduced by placing constraints on the probability that all the tracks originated from the same vertex (vtxprob) and the mass. However tight constraints on the mass and vtxprob can result in some inefficiency. Therefore the constraints are left fairly loose. For the D^0 and D_s mesons the mass windows are 0.07 and 0.04 GeV/c^2 respectively.

 Table 1: Table of decay modes reconstructed for D^0 and D_s^\pm candidates

Mode	Branching Fraction (%) [9]
$D^0 \rightarrow K^- \pi^+$	(3.80 ± 0.09)
$D^0 \rightarrow K^- \pi^+ \pi^0$	(13.1 ± 0.9)
$D^0 \rightarrow K^- \pi^+ \pi^+ \pi^-$	(7.46 ± 0.31)
$D_s^+ \rightarrow \phi \pi^+$	(3.6 ± 0.9)
$D_s^+ \rightarrow K^+ \bar{K}^0(s)$	(3.6 ± 1.1)
$D_s^+ \rightarrow \bar{K}^*(892)^0 K^+$	(3.3 ± 0.9)

6 Tagging

A typical time dependent CP violation analysis requires three steps. First B_{CP} is reconstructed; secondly Δz is measured; thirdly the flavor of the tag B meson is identified. This last step is called tagging and is performed by a software package called BTagger. BTagger assumes the CP eigenstate to have been fully reconstructed and analyzes the remaining decay products to determine the flavor of the tag B. There are a number of physical processes which yield flavor information about the tag B. Each of these physical processes is analyzed by a specific subprogram, or subtagger, of BTagger. Each physical process produces decay products with attributes which are indicative of the flavor of the tag B. In order to exploit correlations between the values of various attributes of the decay products a neural network(See appendix) is used with these attributes as inputs. Based on these inputs the neural network for each subtagger then returns a value between negative one and one, corresponding to the likelihood that the tag B was an antiparticle or a particle respectively. If the subtagger is unable to determine the flavor, it returns zero. The outputs of all the neural networks of the subtaggers are then used as inputs for the BTagger neural network. The output of BTagger is the probability that the tag B is a particle or antiparticle.

6.1 The D^0 and D_s Subtaggers

If a D^0 or a D_s^\pm meson is produced in the decay of the tag B meson, then the flavor of the D^0 or the charge of the D_s^\pm will carry flavor information about the tag B meson. In general a D_s^+ (\bar{D}^0) corresponds to a B^0 , while a D_s^- (D^0) corresponds to a \bar{B}^0 although in a very small fraction of the neutral B decays, 14% of the total number of reconstructed decays, $B^0 \rightarrow D^0 X$ ($\bar{B}^0 \rightarrow \bar{D}^0 X$). The primary function of both the D^0 and D_s^\pm subtaggers is to determine if the reconstructed D^0 or D_s^\pm candidate has been correctly reconstructed. More than half of all reconstructed D^0 and D_s^\pm candidates are not actually D^0 's or D_s^\pm 's respectively, and a falsely reconstructed candidate may carry incorrect flavor information. The neural networks for the D^0 and D_s subtaggers analyze a number of attributes of the decay products of the D^0 or D_s decays. These variables are discussed in the following sections.

6.2 Input Variables for the D^0 Subtagger

The D^0 neural network takes five input variables which are described below.

- *npi* records the number of pions produced in the decay from which the D^0 candidate was reconstructed. $npi=1$ corresponds to $D^0 \rightarrow K^-\pi^+$, $npi=2$ to $D^0 \rightarrow K^-\pi^+\pi^0$ and $npi=3$ to $D^0 \rightarrow K^-\pi^+\pi^+\pi^-$. Of these modes $D^0 \rightarrow K^-\pi^+$ is by far the cleanest with a purity of $\sim 50\%$. For the other decay modes, $D^0 \rightarrow K^-\pi^+\pi^0$ and $D^0 \rightarrow K^-\pi^+\pi^+\pi^-$ the purities are $\sim 19\%$ and $\sim 15\%$ respectively.
- *ktag* is the likelihood that the kaon candidate produced in the decay of the D^0 candidate is actually a kaon. *ktag* is based on the PID information produced by the reconstruction software.
- *pstar* is the momentum of the D^0 candidate in the rest frame of the $\Upsilon(4S)$.
- *mass* is the relativistic invariant mass reconstructed for the D^0 candidate measured in GeV/c^2 . The actual mass of the the D^0 meson is around $1.867 GeV/c^2$.
- *vxprob* is the probability computed by the reconstruction software that all the decay products from the D^0 candidate came from the same vertex.

In addition although it is not actually an input to the neural network, *kq*, the charge of the kaon produced in the decay of the neutral D meson, is used by the subtagger to determine the flavor of the D meson. A positively charged kaon corresponds to a D^0 while a negatively charged kaon corresponds to the antiparticle.

6.3 Input Variables for the D_s^\pm Subtagger

The D_s^\pm neural network has the following six inputs:

- *chanid* records the decay mode from which the D_s candidate was reconstructed. A *chanid* of 1 corresponds to $D_s^+ \rightarrow \phi\pi^+$, 3 corresponds to $D_s^+ \rightarrow K^+\bar{K}^0(s)$, and 4 corresponds to $D_s^+ \rightarrow \bar{K}^*(892)^0K^+$. The purities are 34%, 10% and 14% respectively with $D_s^+ \rightarrow \phi\pi^+$ having the lowest background.
- *costheta* is the cosine of the angle between the momentum vector of the D_s candidate and the z axis in the lab frame. In the lab frame, the momentum vector of the $\Upsilon(4S)$ and therefore the momentum vector of the tag B points

along the z axis. So, $\cos\theta$ is effectively the cosine of the angle between the momentum vector of the tag B and the momentum vector of the D_s candidate.

- $vtxprob$ is the probability computed by the reconstruction software that all the decay products from the D_s candidate came from the same vertex.
- $pstar$ is the momentum of the D_s candidate in the rest frame of the $\Upsilon(4S)$.
- $mass$ is the relativistic invariant mass reconstructed for the D_s candidate measured in GeV/c^2 . The actual mass of the the D_s meson is around 1.968 GeV.
- $d1massdif$ is the difference between the mass of the reconstructed ϕ , $\bar{K}^0(s)$, and $\bar{K}^*(892)^0$ particles in the final state and their actual masses.

For both the D^0 and the D_s subtaggers, $pstar$ and $mass$ are the most powerful variables in differentiating between signal and background.

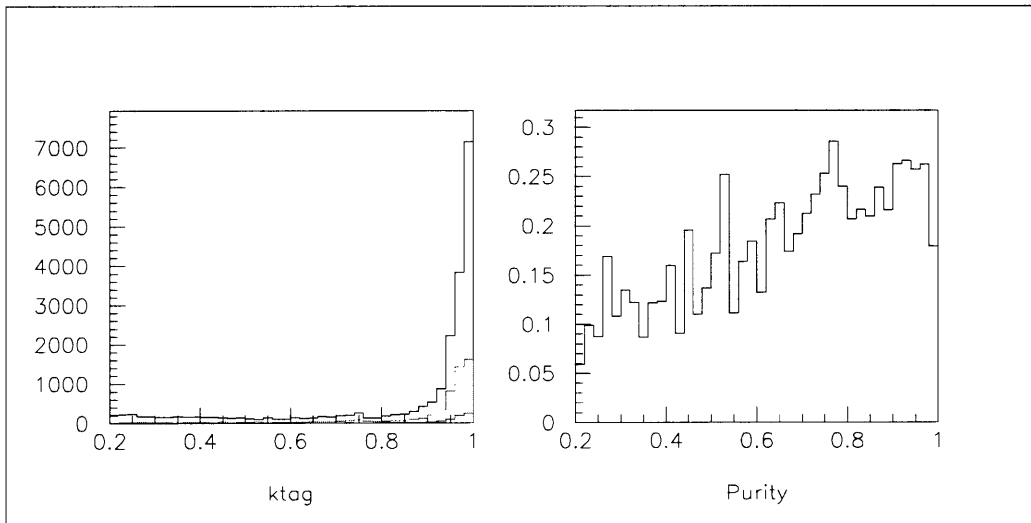


Figure 3: Left: Distribution of k_{tag} for signal from right sign $B^0 \rightarrow \bar{D}^0 X$ in red; for background from wrong sign $B^0 \rightarrow D^0 X$ in blue and for combinatorial background in black. Right: purity ($\frac{signal}{background+signal}$) versus k_{tag} .

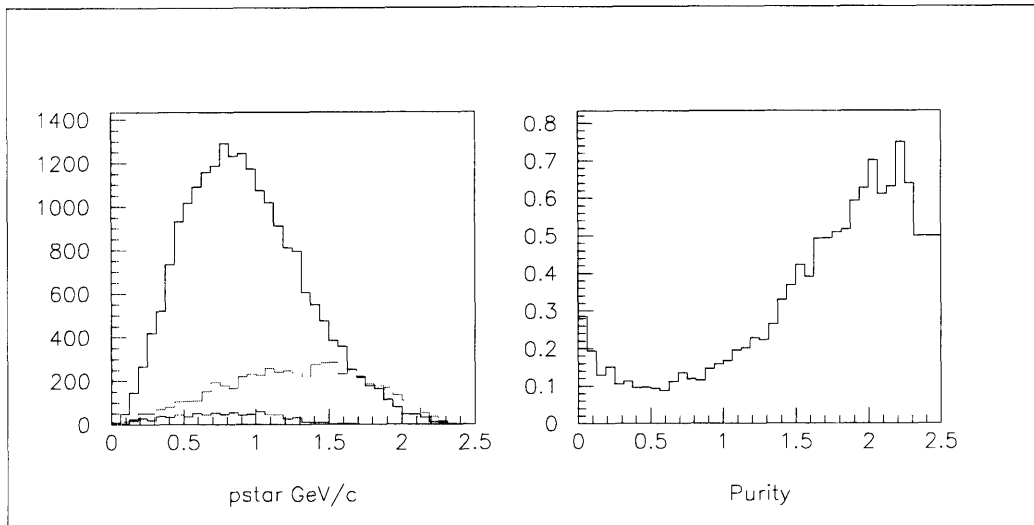


Figure 4: Left: Distribution of $pstar$ for signal from right sign $B^0 \rightarrow \bar{D}^0 X$ in red; for background from wrong sign $B^0 \rightarrow D^0 X$ in blue and for combinatorial background in black. Right: purity $\left(\frac{signal}{background+signal}\right)$ versus $pstar$.

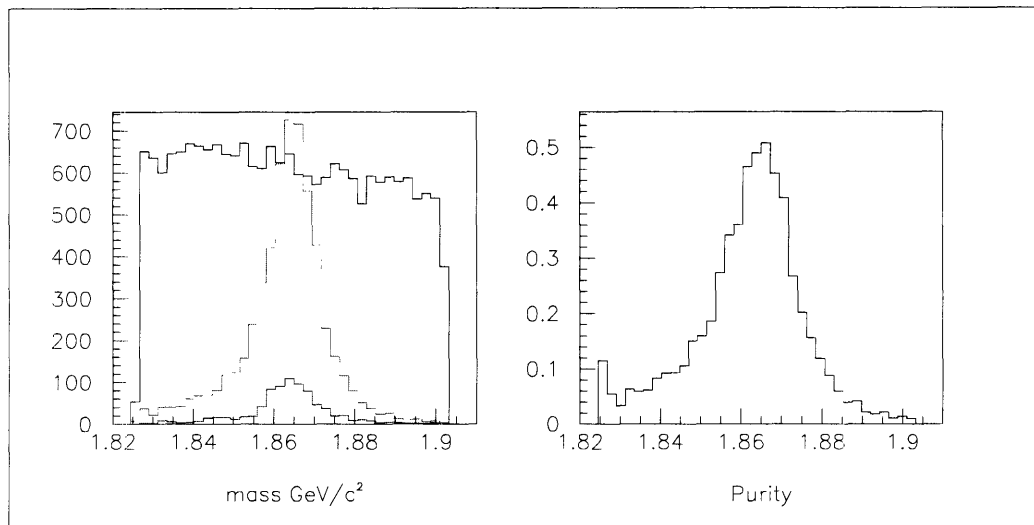


Figure 5: Left: Distribution of the invariant $mass$ for signal from right sign $B^0 \rightarrow \bar{D}^0 X$ in red; for background from wrong sign $B^0 \rightarrow D^0 X$ in blue and for combinatorial background in black. Right: purity $\left(\frac{signal}{background+signal}\right)$ versus $mass$.

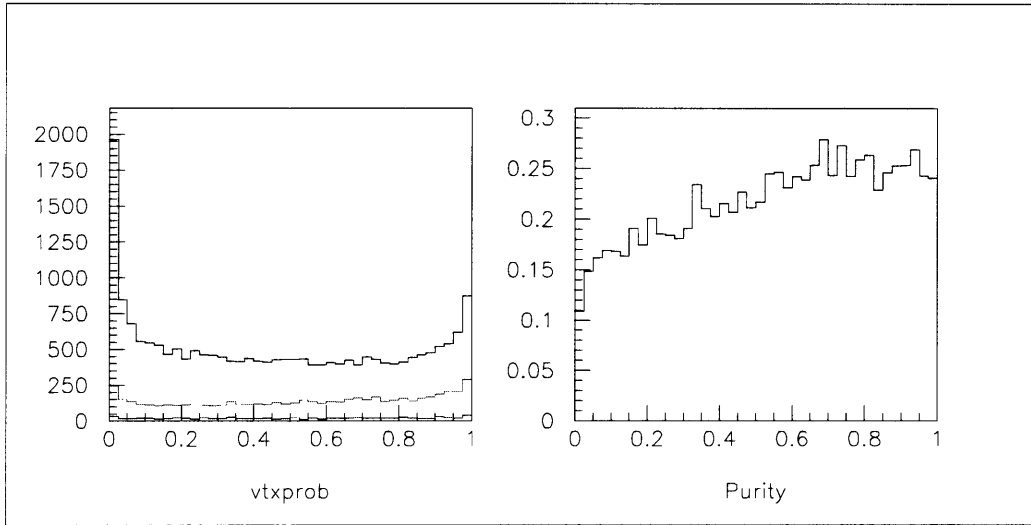


Figure 6: Left: Distribution of $vtxprob$ for signal from right sign $B^0 \rightarrow \bar{D}^0 X$ in red; for background from wrong sign $B^0 \rightarrow D^0 X$ in blue and for combinatorial background in black. Right: purity $\left(\frac{signal}{background+signal}\right)$ versus $vtxprob$.

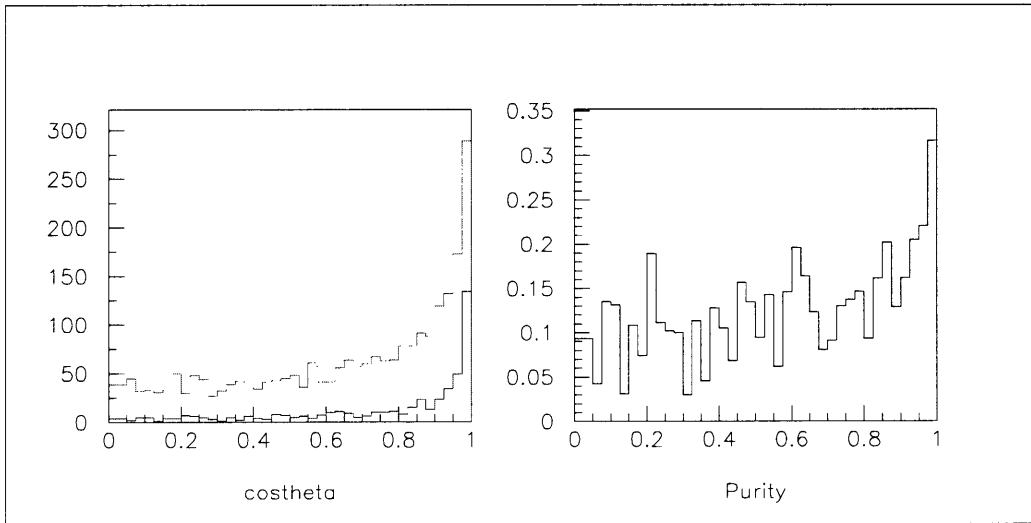


Figure 7: Left: Distribution of $costheta$ for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity $\left(\frac{signal}{background+signal}\right)$ versus $costheta$.

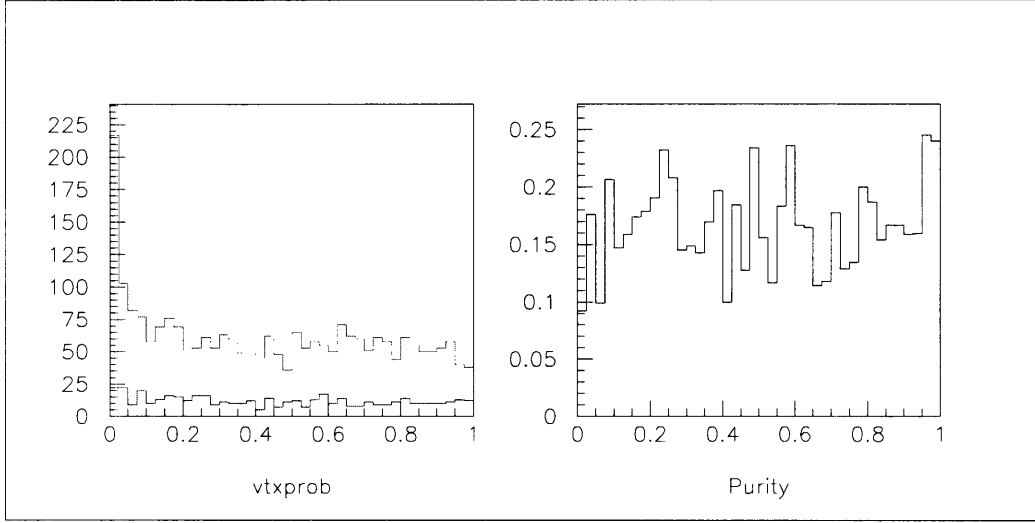


Figure 8: Left: Distribution of $vtxprob$ for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity ($\frac{signal}{background+signal}$) versus $vtxprob$.

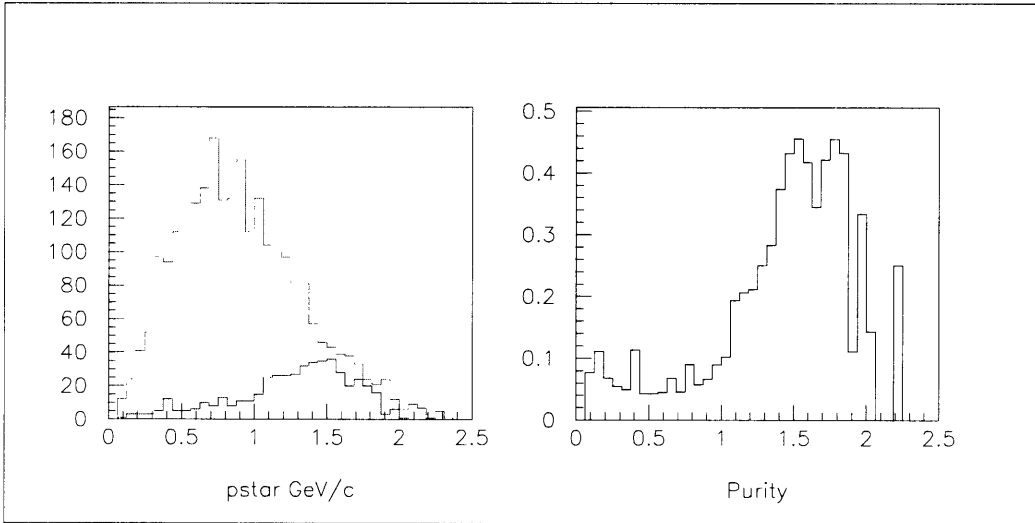


Figure 9: Left: Distribution of $pstar$ for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity ($\frac{signal}{background+signal}$) versus $pstar$.

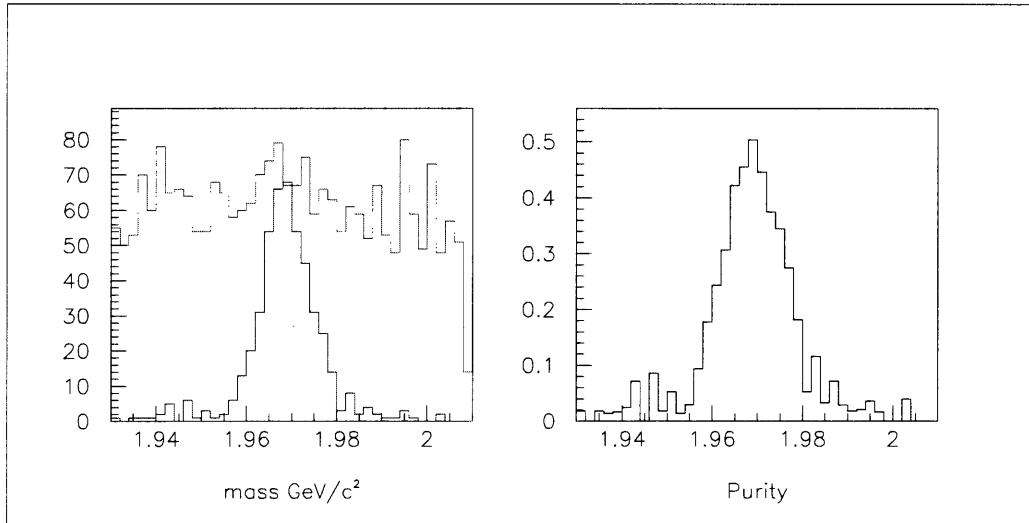


Figure 10: Left: Distribution of the invariant $mass$ for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity ($\frac{signal}{background+signal}$) versus $mass$.

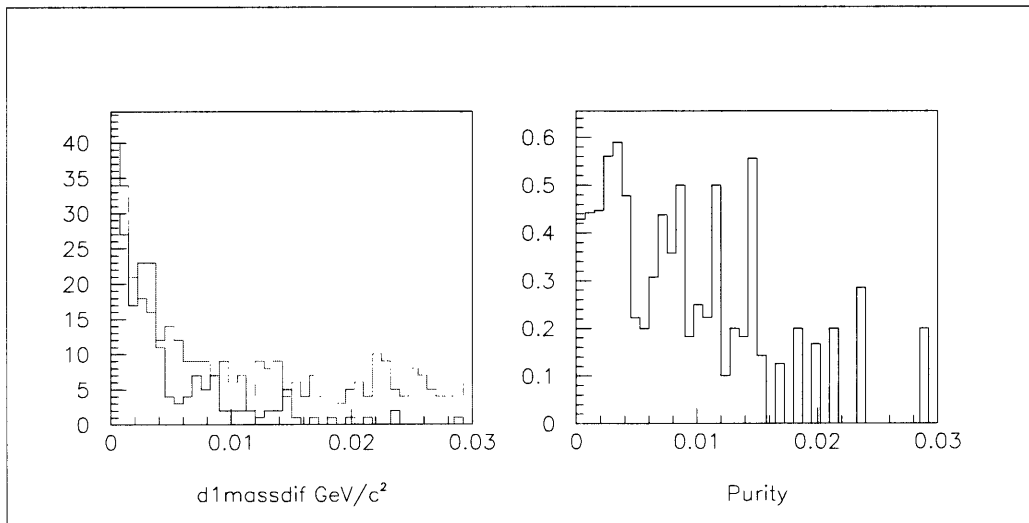


Figure 11: Left: Distribution of $d1massdif$ in $D_s^+ \rightarrow \phi\pi^+$ decays for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity versus $d1massdif$.

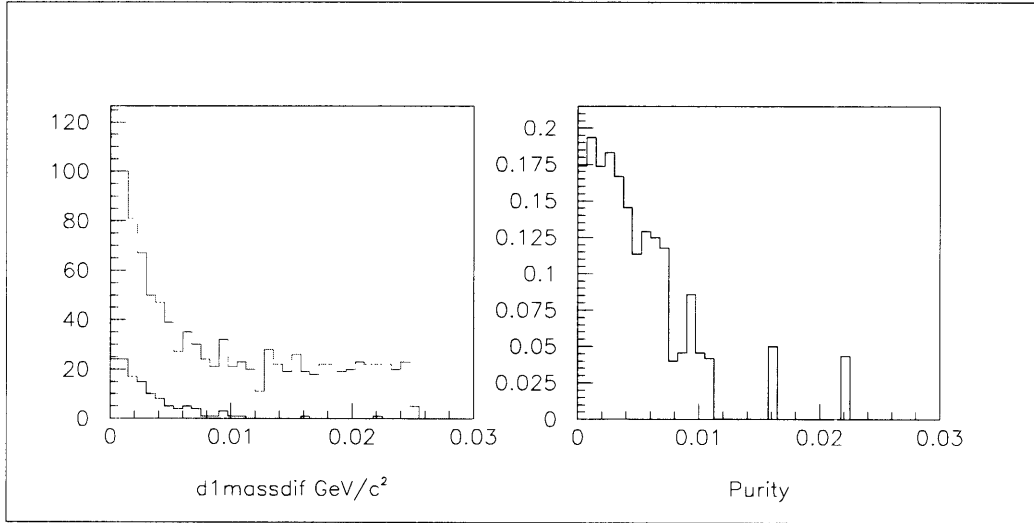


Figure 12: Left: Distribution of $d1massdif$ in $D_s^+ \rightarrow K^+ \bar{K}^0(s)$ decays for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity versus $d1massdif$.

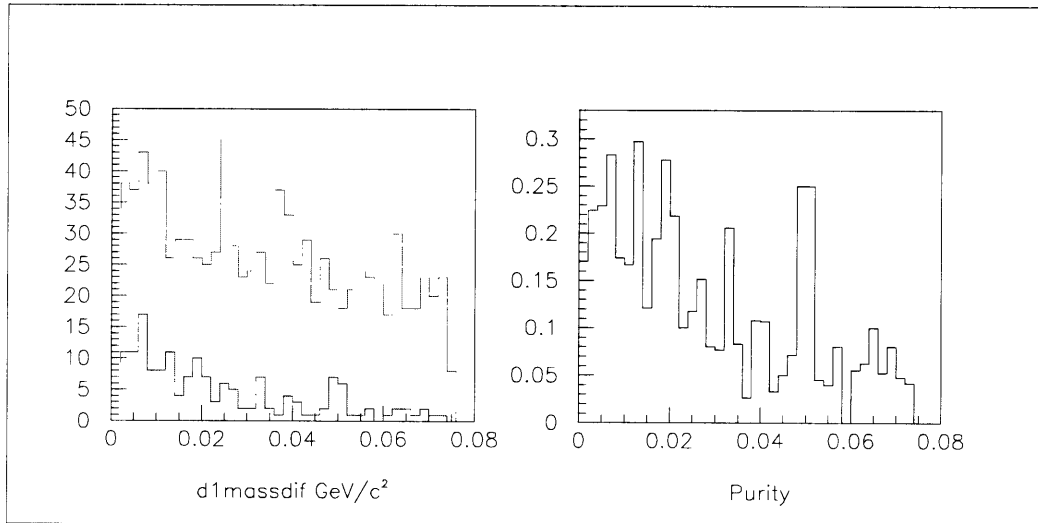


Figure 13: Left: Distribution of $d1massdif$ in $D_s^+ \rightarrow \bar{K}^*(892)^0 K^+$ decays for signal from $B^0 \rightarrow D_s^+ X$ in blue and for combinatorial background in red. Right: purity versus $d1massdif$.

6.4 The Effective Tagging Efficiency (Q)

The amplitude of the time-dependent CP asymmetry $A_{CP}^{observed}(t)$ measured in the experiment is related to the theoretical $A_{CP}^{theory}(t)$ defined in equation (9) by the formula:

$$A_{CP}^{observed}(t) = \sqrt{Q} A_{CP}^{theory}(t) \quad (10)$$

Q is the ‘‘effective tagging efficiency’’, defined as:

$$Q = \epsilon^{tag}(1 - 2w)^2 \quad (11)$$

where ϵ^{tag} is the probability of tagging an event as a B^0 or \bar{B}^0 and w is the probability of tagging a B^0 as a \bar{B}^0 or vice-versa. To make better use of the information contained in the data the definition above is slightly modified and Q is calculated separately in bins of the output of the tagger and separately for B^0 and \bar{B}^0 as explained below.

For each tag B flavor the output of the tagger or subtagger neural network is plotted in a 40 bin histogram. N is the total number of decay events for the tag B flavor. For the histogram of the output of BTagger, the total number of entries is N . However in the output of the subtagger neural networks many of the events may not be plotted if the physical process analyzed by the subtagger was not present in these events. In this case, N will exceed the total number of events plotted. n_i^{RS} is the number of correctly tagged events in the i 'th bin, and n_{-i}^{WS} is the number of incorrectly tagged events in the corresponding bin of opposite sign. For bins where n_{-i}^{WS} is greater than n_i^{RS} both n_i^{RS} and n_{-i}^{WS} are reset to 0.

$$\epsilon_i = \frac{n_i^{RS} + n_{-i}^{WS}}{N} \quad (12)$$

$$\omega_i = \frac{n_{-i}^{WS}}{n_i^{RS} + n_{-i}^{WS}} \quad (13)$$

$$Q_i^{B^0(\bar{B}^0)} = \epsilon_i \times (1 - 2\omega_i)^2 \quad (14)$$

For simulated decays of both tag B flavors and for each pair of bins $Q_i^{B^0(\bar{B}^0)}$ is calculated. $Q_i^{B^0}$ is calculated for events involving B^0 decays, and $Q_i^{\bar{B}^0}$ is calculated for events involving \bar{B}^0 decays.

$$Q_i = \frac{1}{2}(Q_i^{B^0} + Q_i^{\bar{B}^0}) \quad (15)$$

$$Q = 100 \times \sum_{i=1}^{20} Q_i \quad (16)$$

For each subtagger, Q is affected by the inputs used and by the training of the neural network. The neural networks are trained using MC (Monte Carlo) generated

events. The MC produces 50,000 events where $B_{CP} = B^0 \rightarrow \pi^+\pi^-$ with mixing turned off so that the tag B for these events is a \bar{B} when it decays. Likewise another 50,000 events are produced where $B_{CP} = \bar{B}^0 \rightarrow \pi^+\pi^-$ and $B_{tag} = B^0$. The neural network training algorithm(See appendix: A.2) is run on 50,000 of the MC generated events - 25,000 for each tag B flavor. A step size is first set and then a series of cycles is run until the *MSE* (Mean Squared Error) reaches a minimum and levels out. This usually takes around two hundred cycles. Then the steps size is reset to a smaller value and another series of cycles is run and so on. Typically the steps range over the values: 0.5, 0.1, 0.05 ,0.01, 0.005, to 0.001.

To test the neural network after it is trained the tagging software is run on the MC generated events which were not used in training, and Q is calculated. Changing the step size or the number of cycles per step has little affect on Q. In general such alterations result in a change in Q of only a few hundredths.

6.5 The Results

6.5.1 Correlation Between Inputs and Output for the D^0 and D_s Subtaggers

If the neural networks for the D^0 and D_s subtaggers are properly trained, then there should be a correlation between high tag values and high purities in the input variables. This in fact holds for the D^0 and D_s subtagger neural networks as the figures 14 - 20.

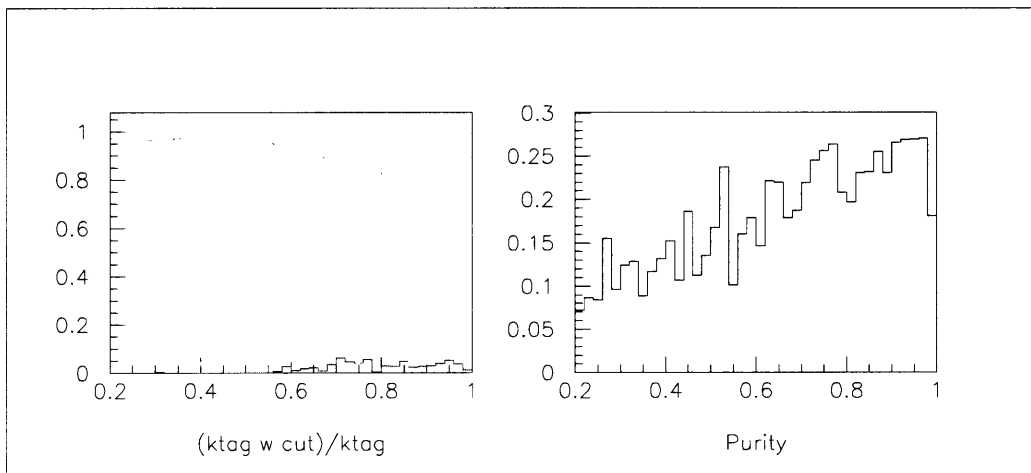


Figure 14: For the D^0 subtagger the distribution of $ktag$ values for candidates with output values within a given interval divided by the $ktag$ values with no cut on the output; blue: output > 0.7 , cyan: $0.7 > \text{output} > 0.4$, green: $0.4 > \text{output}$. For reference the purity versus $ktag$ is graphed in black to the right.

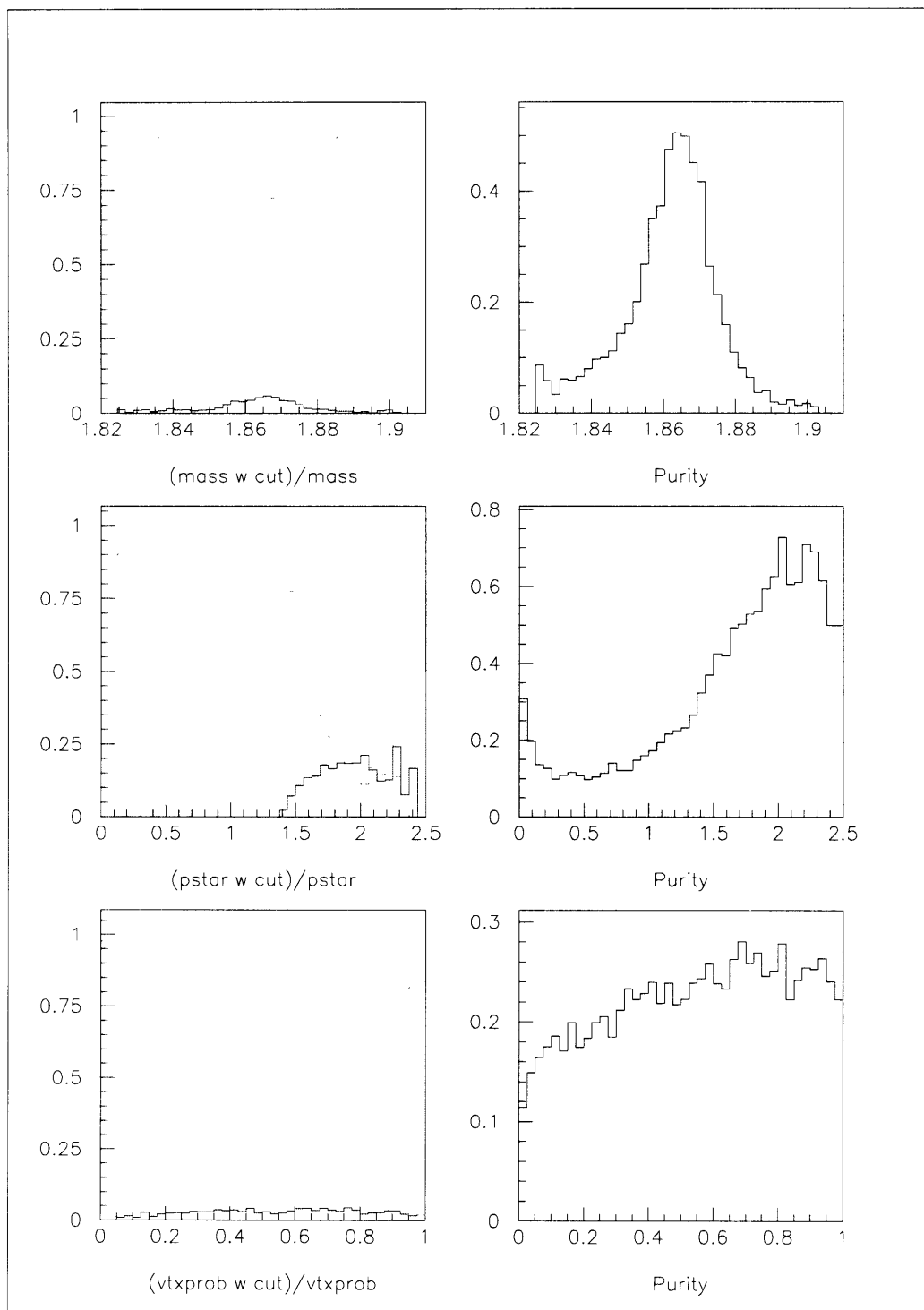


Figure 15: For the D^0 subtagger the distribution of input variable values for candidates with output values within a given interval divided by the input variable values with no cut on the output; blue: output > 0.7 , cyan: $0.7 > \text{output} > 0.4$, green: $0.4 > \text{output}$. For reference the purity versus each input variable is graphed in black to the right.

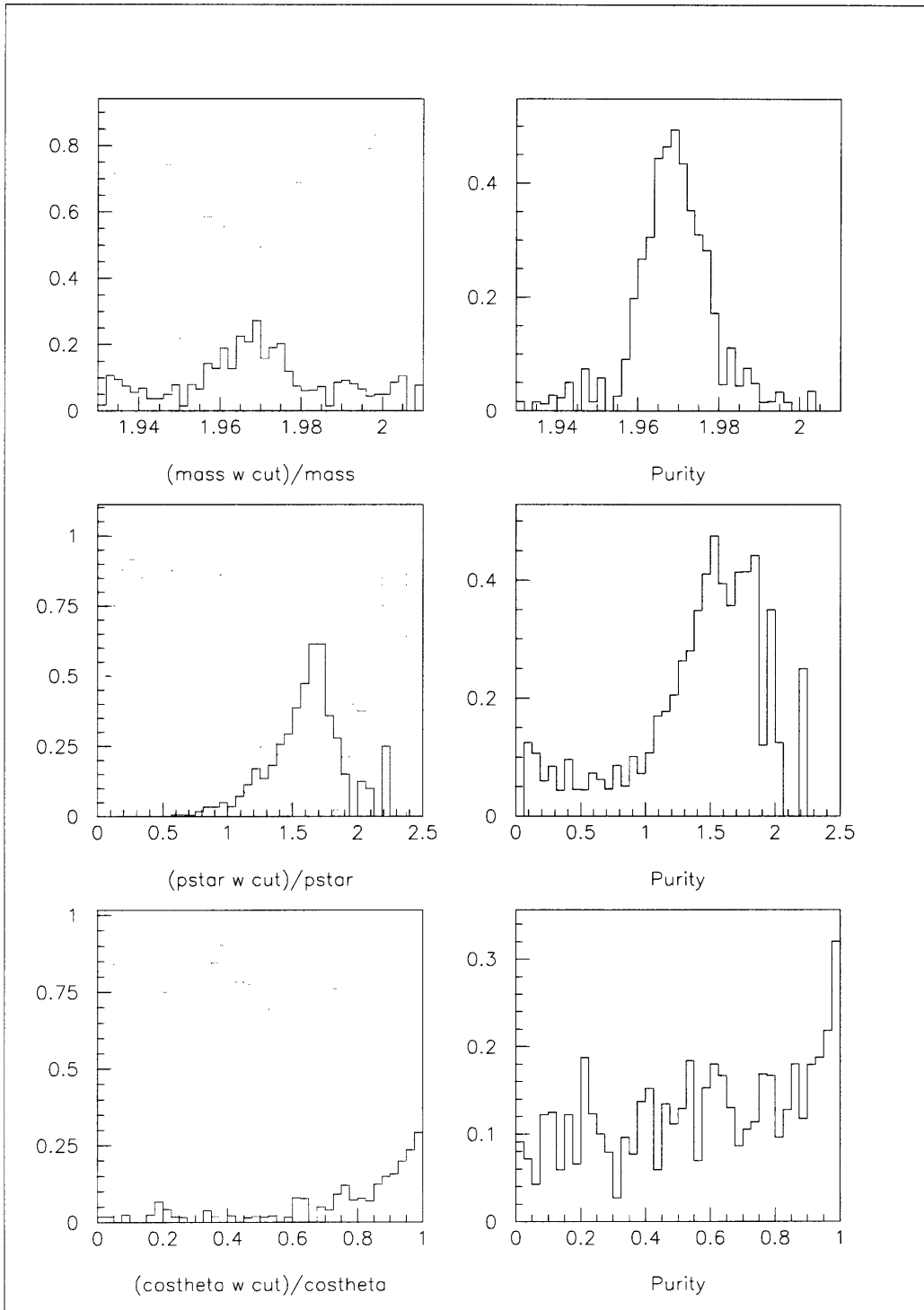


Figure 16: For the D_s subtagger the distribution of input variable values for candidates with output values within a given interval divided by the input variable values with no cut on the output; blue: output > 0.4 , cyan: $0.4 > \text{output} > 0.2$, green: $0.2 > \text{output}$. For reference the purity versus each input variable is graphed in black to the right.

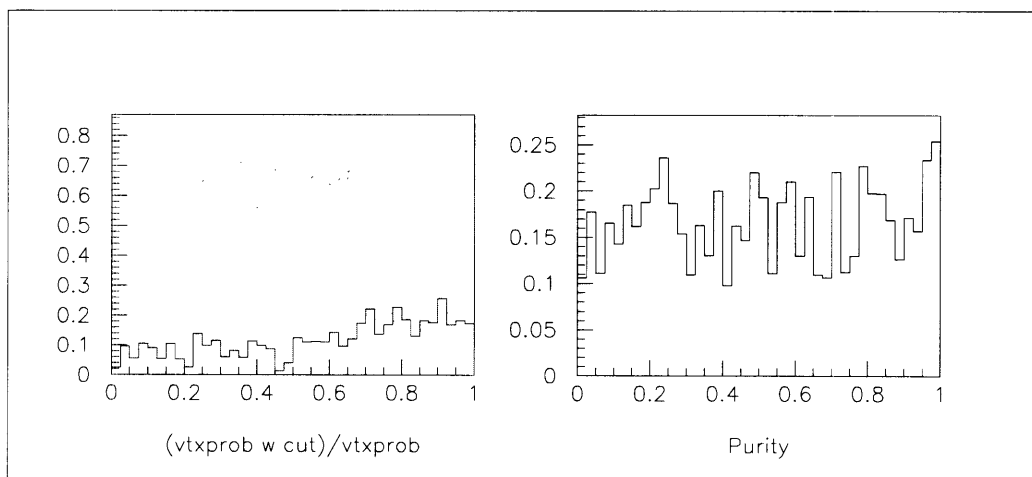


Figure 17: For the D_s subtagger the distribution of $vtxprob$ values for candidates with output values within a given interval divided by the $vtxprob$ values with no cut on the output; blue: output > 0.4, cyan: $0.4 > \text{output} > 0.2$, green: $0.2 > \text{output}$. For reference the purity versus $vtxprob$ is graphed in black to the right.

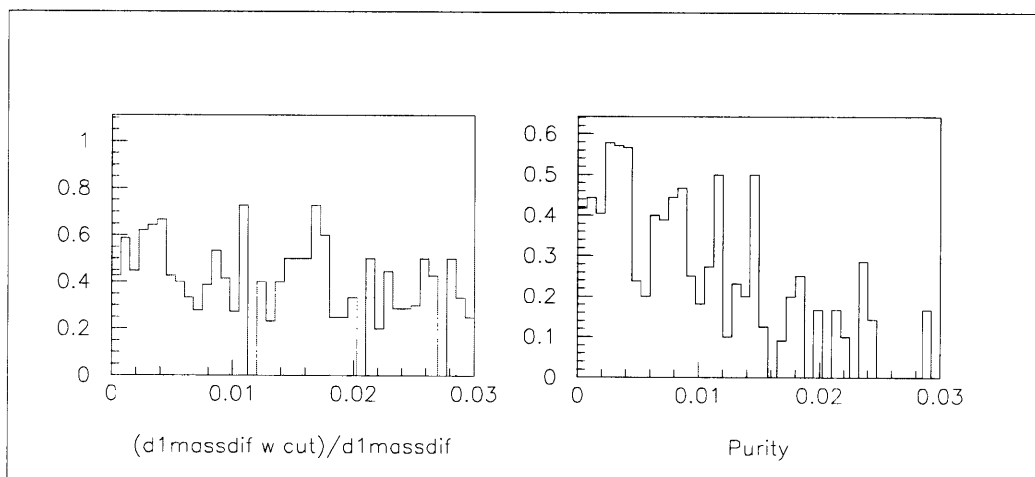


Figure 18: For the D_s subtagger the distribution of $d1massdif$ values for candidates from $D_s^+ \rightarrow \phi\pi^+$ decays with output values within a given interval divided by the $d1massdif$ values with no cut on the output; blue: output > 0.4, cyan: $0.4 > \text{output} > 0.2$, green: $0.2 > \text{output}$. For reference the purity versus $d1massdif$ is graphed in black to the right.

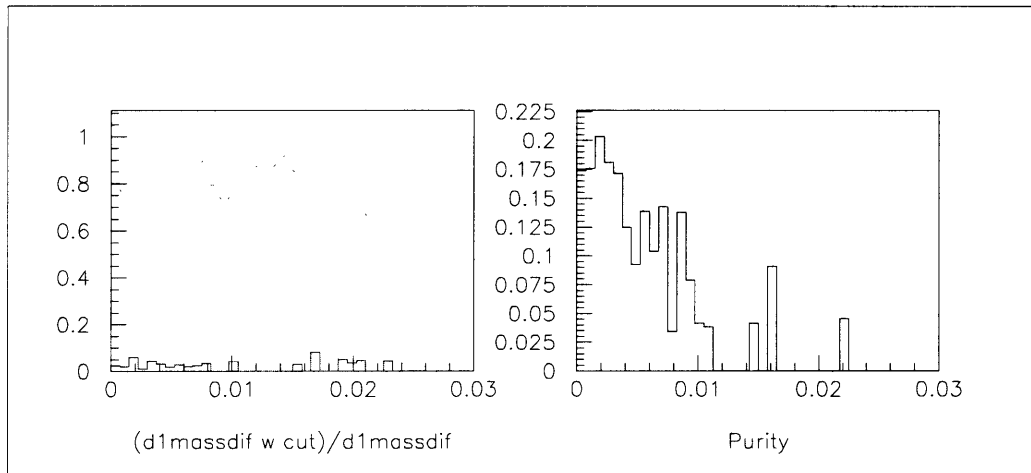


Figure 19: For the D_s subtagger the distribution of $d1massdif$ values for candidates from $D_s^+ \rightarrow K^+ \bar{K}^0(s)$ decays with output values within a given interval divided by the $d1massdif$ values with no cut on the output; blue: output > 0.4 , cyan: $0.4 > \text{output} > 0.2$, green: $0.2 > \text{output}$. For reference the purity versus $d1massdif$ is graphed in black to the right.

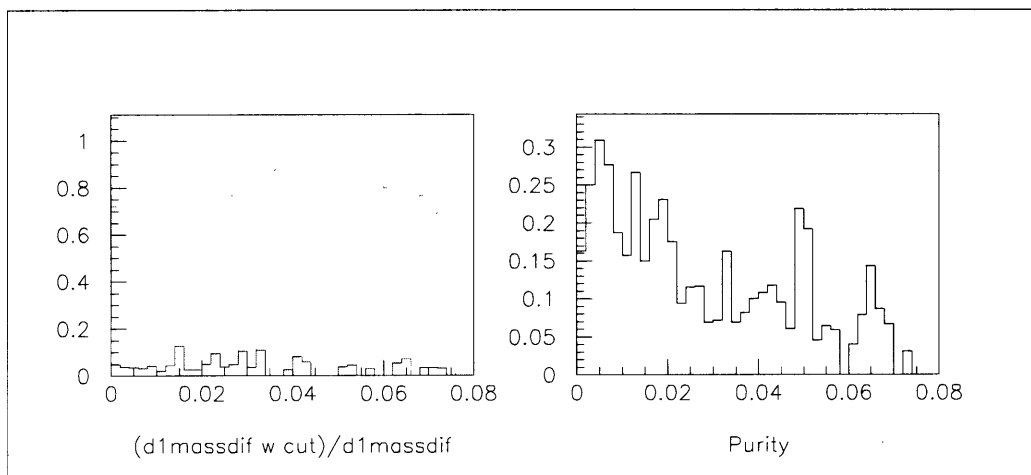


Figure 20: For the D_s subtagger the distribution of $d1massdif$ values for candidates from $D_s^+ \rightarrow \bar{K}^*(892)^0 K^+$ decays with output values within a given interval divided by the $d1massdif$ values with no cut on the output; blue: output > 0.4 , cyan: $0.4 > \text{output} > 0.2$, green: $0.2 > \text{output}$. For reference the purity versus $d1massdif$ is graphed in black to the right.

6.5.2 D^0 Subtagger Performance

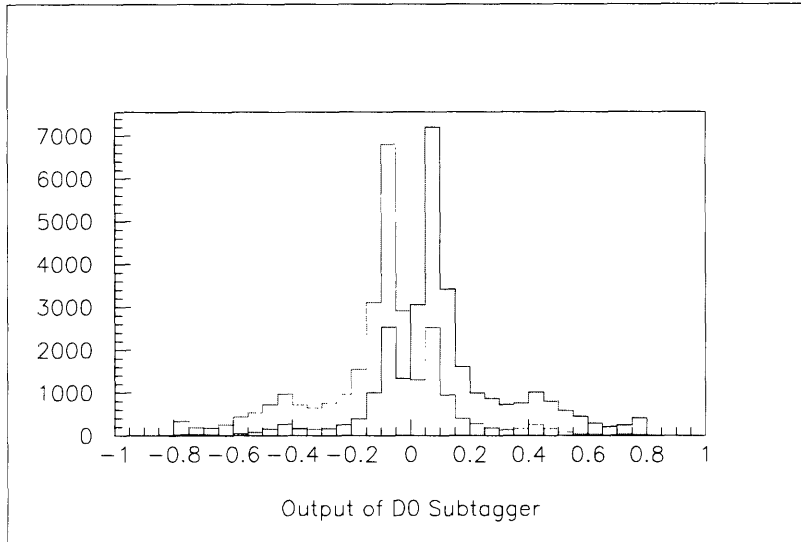


Figure 21: Distribution of the output of the neural network for the D^0 subtagger for candidates from B^0 decays in blue and \bar{B}^0 decays in red.

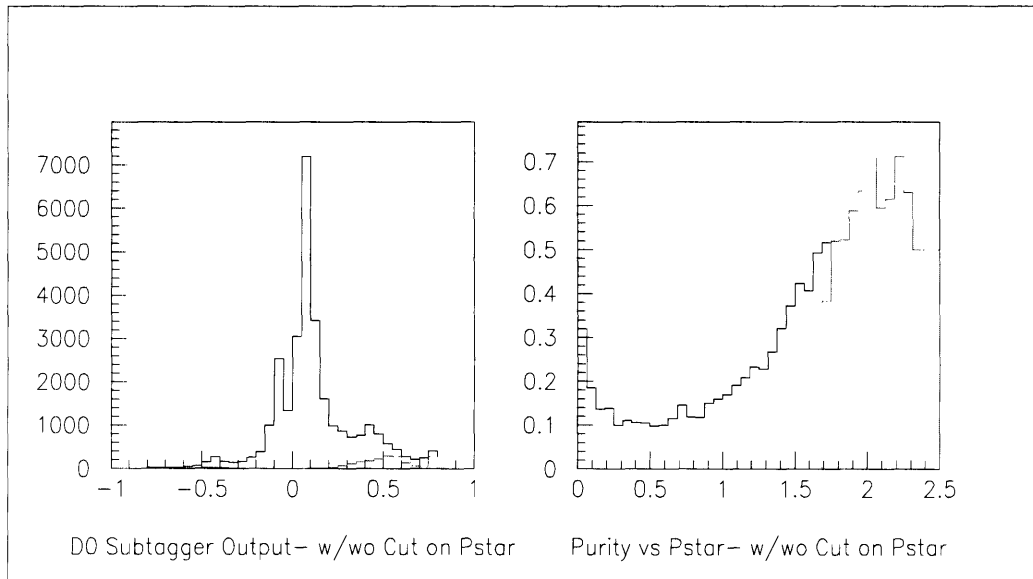


Figure 22: Left: Distribution of the output from the D^0 subtagger for B^0 decays with $p_{star} > 1.7$ in red and $p_{star} > 0$ in black. Right: Purity vs. p_{star} with same color coding.

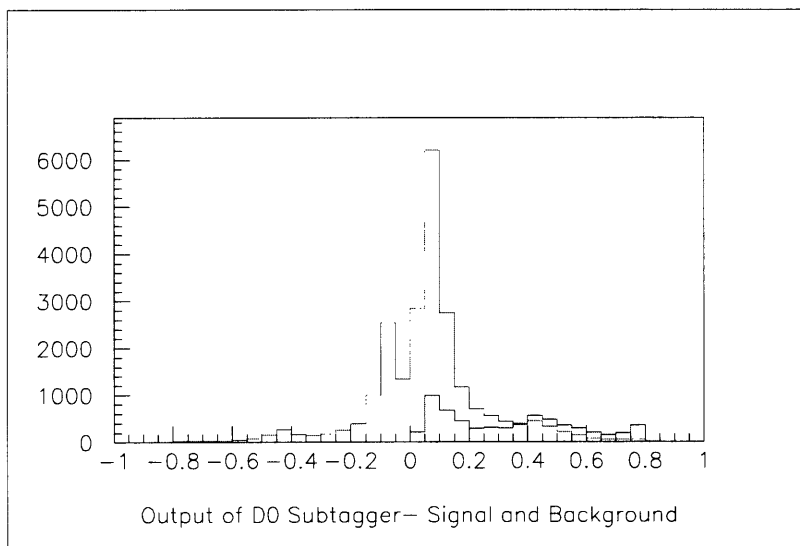


Figure 23: Distribution of the output from D^0 subtagger for candidates from B^0 decays. Signal is graphed in blue and background in red.

The D^0 neural network assigns to a candidate an absolute tag value based on the probability that it is signal and then multiplies the absolute tag value by the kaon charge. The D^0 subtagger performs best if its neural network cleanly separates real D^0 candidates from combinatorial background. However only the candidates with the highest $pstar$ values are cleanly separated from the background(See figure 22). These candidates are concentrated near -0.8 and 0.8 in the output. About 3.4% of the tag B mesons decay to a real neutral D meson candidate which is reconstructed. This is the upper limit of the contribution to Q due to true D^0 candidates irregardless of neural network performance. The actual Q for the D^0 subtagger is about 4.5%. This high Q is not due however to signal but rather to right-sign background(See figure 23). Since the sign of the tag is based upon the sign of the kaon, so long as the kaon daughter of the D^0 candidate has the correct sign, the tag will have the right sign irregardless of whether the D^0 candidate was correctly reconstructed. However when the D^0 subtagger is added to BTagger, the right-sign background in the D^0 subtagger does not contribute to the net effectiveness of BTagger since events which contribute to the background , such as $B^0 \rightarrow K^+ X$ decays, are already tagged more effectively by other subtaggers.

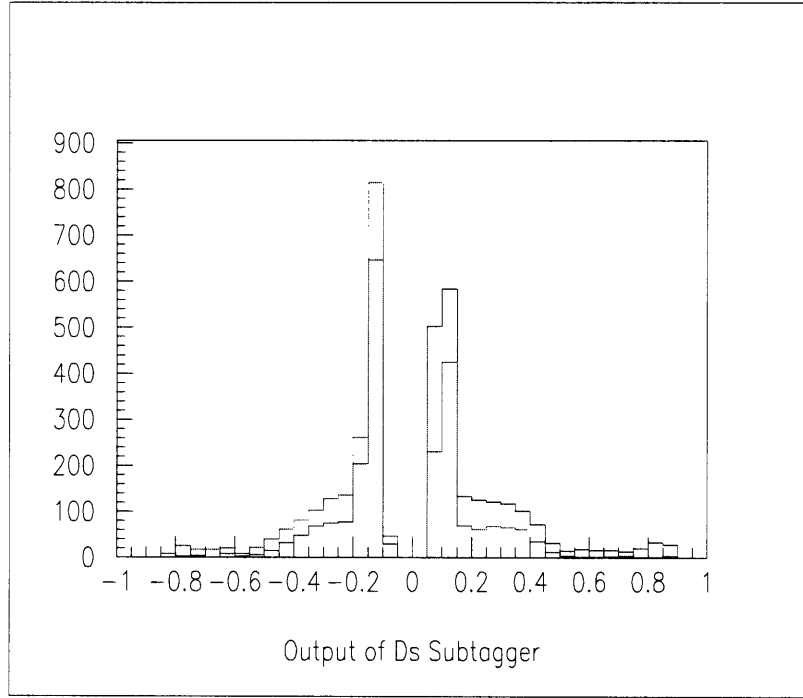


Figure 24: Distribution of the output of the neural network for the D_s subtagger for candidates from B^0 decays in blue and \bar{B}^0 decays in red.

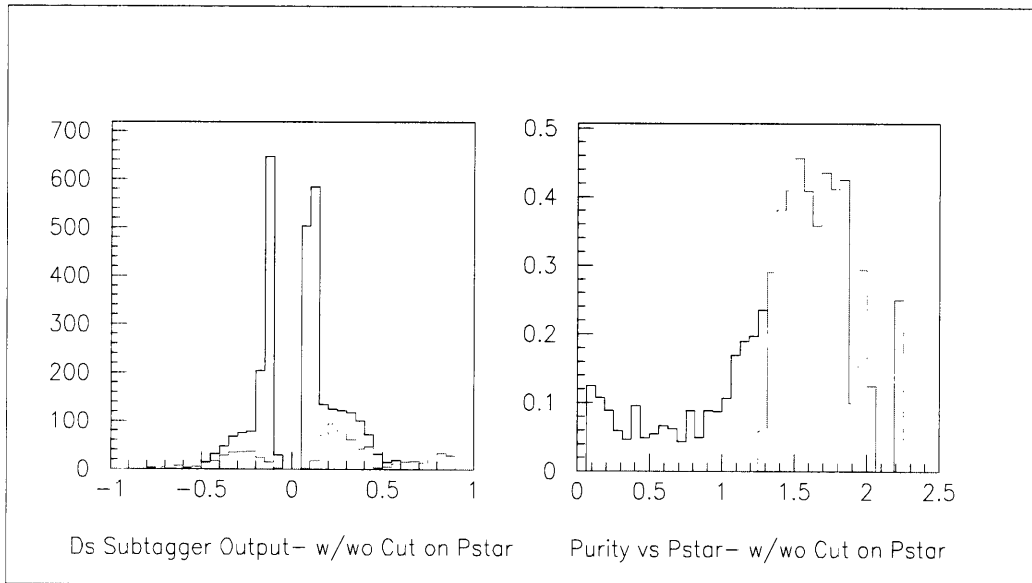


Figure 25: Left: Distribution of the output from the D_s subtagger for B^0 decays with $p_{star} > 1.3$ in red and $p_{star} > 0$ in black. Right: Purity vs. p_{star} with same color coding.

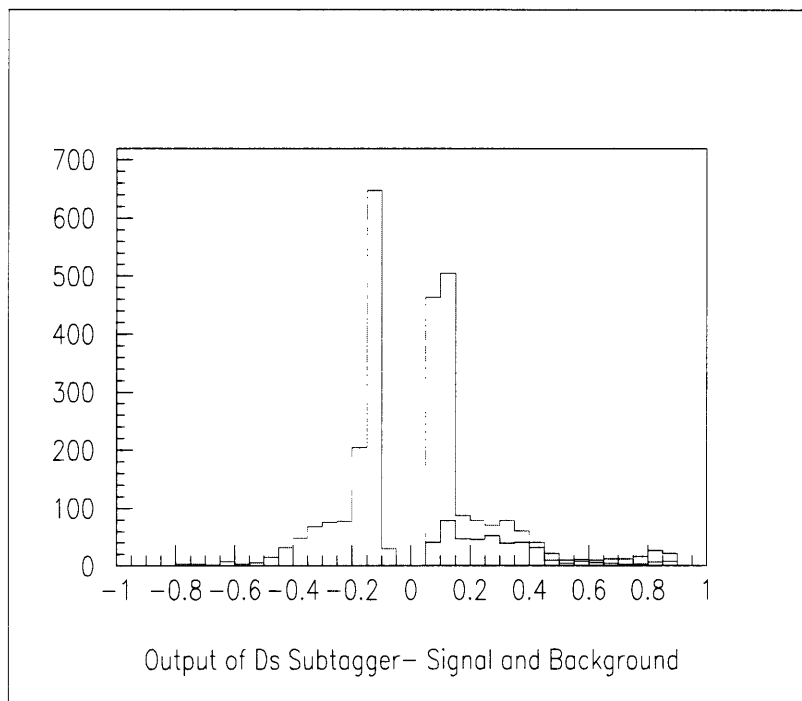


Figure 26: Distribution of the output from D_s subtagger for candidates from B^0 decays. Signal is graphed in blue and background in red.

6.5.3 D_s Subtagger Performance

The D_s neural network assigns to D_s candidates an absolute tag value based on the probability that they are correctly reconstructed and then multiplies their absolute tag value by their charge. The D_s subtagger performs best if its neural network cleanly separates real D_s candidates from combinatorial background. However that is not the case; except for a very small fraction of the candidates, the signal to background ratio is worse than 50% even for candidates with high $pstar$ values (See figure 25). About 0.29% of the tag B mesons decay to a real D_s candidate which is reconstructed. This is the upper limit of the contribution to Q due to true D_s candidates irregardless of neural network performance. The actual Q for the D_s subtagger is about 0.12- 0.14%.

6.5.4 BTagger Performance

Including either the D^0 or D_s subtagger does not improve the Q of BTagger. When the neural network for BTagger was retrained including either one or both of these subtaggers, the change in Q was consistent with zero. This is due to the low yields of

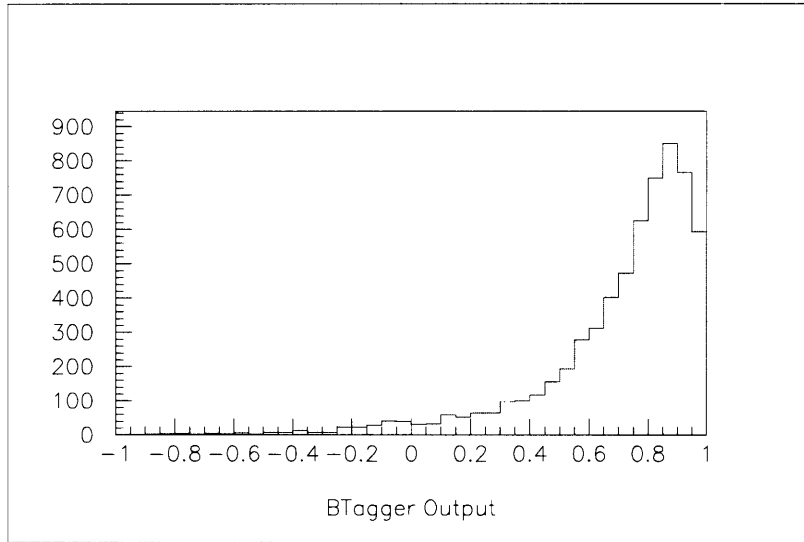


Figure 27: Distribution of the output of BTagger without the D^0 or D_s subtaggers included for $B^0 \rightarrow \bar{D}^0 X$ decays where the \bar{D}^0 was correctly reconstructed.

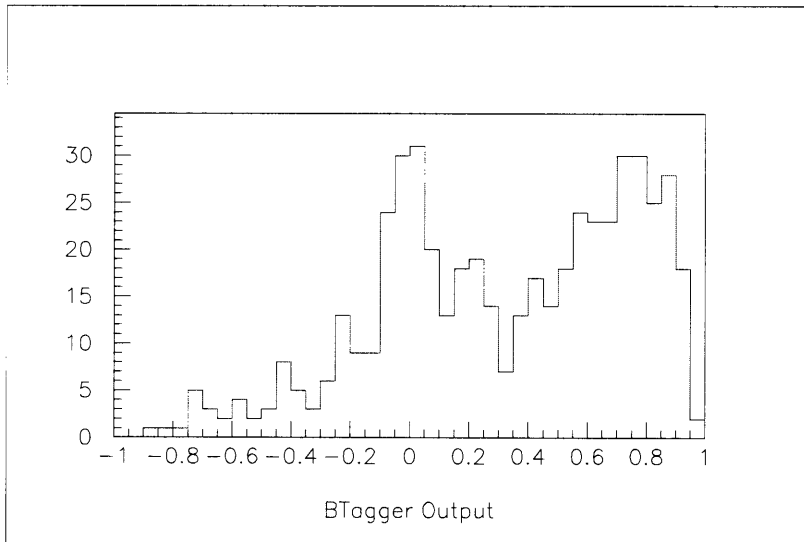


Figure 28: Distribution of the output of BTagger without the D^0 or D_s subtaggers included for $B^0 \rightarrow D_s^+ X$ decays where the D_s^+ was correctly reconstructed.

true D^0 or D_s^\pm meson candidates from B decays, only $\sim 3.4\%$ or $\sim 0.29\%$, respectively and the fact that a significant majority of the events producing a D^0 or D_s are already correctly tagged by the other subtaggers, such as the kaon subtagger. Only 0.1% (0.06%) of B decays produce a D^0 (D_s) which is correctly tagged by the subtagger but incorrectly tagged by BTagger even without the D^0 and D_s subtaggers included as inputs to BTagger.

7 Conclusion

BaBar was one of the first two experiments to establish the existence of CP violation in the B meson system, demonstrating that CP violation is not just an anomaly of the neutral kaon system. CP violation has a number of deep implications. It suggests that fundamental interactions between particles are not invariant under time reversal, and also fulfills one of the conditions necessary to explain the existence of the universe as it is needed to explain the small excess of matter created in the Big Bang. In the Standard Model CP violation arises in the quark sector as a result of a complex phase in the CKM matrix. Precise measurements of the parameters governing CP violation can be used to test the Standard Model. The BaBar experiment continues to refine its measurements of these parameters. The BaBar experiment measures the CP violating asymmetry by measuring the decay time distribution to a CP eigenstate separately for B^0 and \bar{B}^0 decays from a $B^0\bar{B}^0$ pair. This measurement is possible since the two B mesons, B_{tag} and B_{CP} , evolve coherently until one of them decays and therefor the flavor of B_{CP} can be obtained by tagging the other meson. The precision of this measurement could be improved by improving the effectiveness of the tagging algorithm. To this algorithm I added the D^0 and D_s subtaggers which estimated the tag of the neutral B meson based on the fact that in general $B^0 \rightarrow D_s^+ X, \bar{D}^0 X$. The Q values obtained for the D^0 and D_s subtagger were $\sim 4.5\%$ and $\sim 0.12 - 0.14\%$ respectively. However adding the subtaggers did not improve the performance of the tagging algorithm due to the very low percentage of B mesons which decayed into either a D^0 or D_s^\pm and the fact that most events involving such decays were already correctly tagged.

8 Acknowledgments

I wish to thank my thesis advisor Professor Sciolla.

References

- [1] Aubert B *et al.* (Babar Collaboration) 2001 *Preprint* hep-ex/0102030
- [2] Aubert B *et al.* (Babar Collaboration) 2001 *Preprint* hep-ex/0105044
- [3] Aubert B *et al.* (Babar Collaboration) 2002 *Preprint* hep-ex/0201020
- [4] Aubert B *et al.* (Babar Collaboration) 2000 *Preprint* hep-ex/0012042
- [5] A. Abashian *et al.* (Belle Collaboration), *Phys. Rev. Lett.* **86**, 2509 (2001).
- [6] Britton, D.I., M.J.Ryan, X.Qu, “Studies of $PbWO_4$ Endcap Crystals Using a Hybrid Photomultiplier Tube,” *CMS Note* 2003/018
- [7] Buckley, James J., Thomas Feuring, *Fuzzy and Neural: Interactions and Applications*, 1999, (New York: Physica-Verlag)
- [8] Christenson, J.H., J.W. Cronin, V.L. Fitch, and R.Turlay, “Evidence for the 2π Decay of the K_2^0 Meson”, *Phys. Rev. Lett.* **13**, 138 (1964).
- [9] Eidelman, S. *et al.*, *Review of Particle Physics*, *Physics Letters B* **592**, 1 (2004)
- [10] Griffiths, David, *Introduction to Elementary Particles*, 1987, (New York: John Wiley & Sons)
- [11] In Publication: Harrison, P.F., M. Morii and G. Sciolla, “B Physics at e^+e^- B Factories,” 2005, Submitted to: *Rep. Prog. Phys.*
- [12] Kobayashi M and Maskawa T 1973 *Prog. Theor. Phys.* **49** 652
- [13] Sather, Eric, “The Mystery of the Matter Asymmetry,” *BeamLine* (Spring/Summer 1996) Vol 26, No.1
- [14] Lange, D., G. Sciolla, M.C. Simani, Y. Zheng, “Tag04: a Improved Tagging Algorithm for CP Violation Studies,” 2004, *BaBar Analysis Document* 729
- [15] Lange, D., G. Sciolla, M.C. Simani, Y. Zheng, “BtgTest and Tag04: Studies Towards an Improved Tagging Algorithm,” 2004, *BaBar Analysis Document* 730
- [16] Lee, T.D., and C.N. Yang, *Phys. Rev.* **104**, 254 (1956).

- [17] Nguyen, Hung T. *et al.*, *A First Course in Fuzzy and Neural Control*, 2003 (Boca Raton, FL: Chapman & Hall/CRC Press)
- [18] Sakharov A D 1967 *Pisma Zh. Eksp. Teor. Fiz.* **5** 32 (*JETP Lett.* **5** 24)
- [19] Wolfenstein L 1983 *Phys. Rev. Lett.* **51** 1945
- [20] Wu, C.S. *et al.*, *Phys. Rev.* **105**, 1413 (1957).
- [21] Yannis Karyotakis and Gautier Hamel de Monchenault, "A Violation of CP Symmetry in B Meson Decays", *Europhysics News* (2002) Vol.33 No. 3.
- [22] Fonte, P., "Applications and New Developments in Resistive Plate Chambers", http://ppd.fnal.gov/experiments/e907/TOF/Alice_RPC_Design.pdf
- [23] <http://lhc-b.web.cern.ch/LHC-B/loi/loi-lhcb/node1.html>
- [24] <http://universe-review.ca/R02-14-CPviolation.htm>
- [25] http://hepwww.physics.yale.edu/www_info/babar/babdesc.html
- [26] <http://www.unifr.ch/physics/3cycle/5.partricleidentification.pdf>
- [27] http://www.slac.stanford.edu/BFROOT/www/Detector/DIRC/dirc_whatsit.html
- [28] <http://www.slac.stanford.edu/BFROOT/www/Detector/Calorimeter/software/geometry.html>

A Neural Networks

A.1 Separating Correlated Inputs

The neural networks are used to analyze data from correlated inputs and based upon the values of these inputs determine which of two possible outcomes the inputs favor. A neural network is composed of layers of nodes. The outputs of the nodes from one layer are the inputs of the nodes in the next layer. Suppose a node accepts the inputs x_1, \dots, x_n . The node's output is $n = f(\sum_{i=1}^n x_i \omega_i - b)$ here f is a function which takes values between 0 and 1 such as the step function or the sigmoid function $\frac{1}{1+e^{-x}}$. To simplify notation, set $x_0 = 1$ and $\omega_0 = -b$ then $n = f(\sum_{i=0}^n x_i \omega_i)$. Suppose set A is the set of data points corresponding to outcome 1 and set B is the set of data points corresponding to outcome 2. Furthermore suppose that these sets are linearly separable (See figure 29).

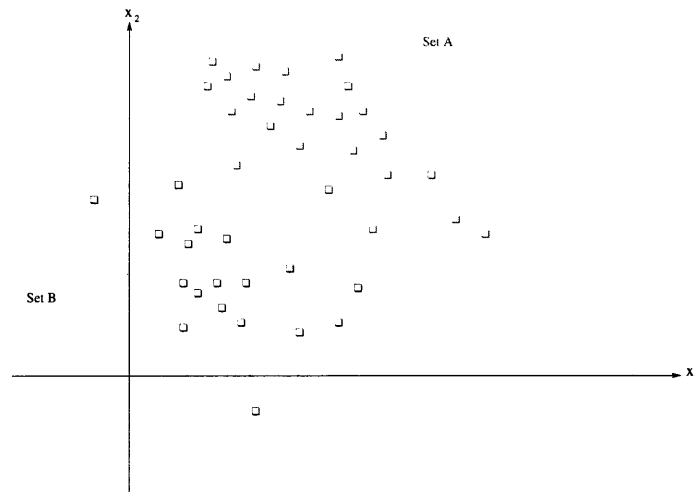


Figure 29: Graph of linearly separable sets A and B .

The node n sends every data point on one side of a hyperplane in \mathcal{R}^n to 1 and every data point on the other side to 0. The location of this hyperplane is determined by the ω_i 's, the weights. So, for the proper choice of ω_i 's the hyperplane of the node can be placed between the sets A and B in such a way that n sends all points in A to 1 and all points in B to 0 (See figure 30).

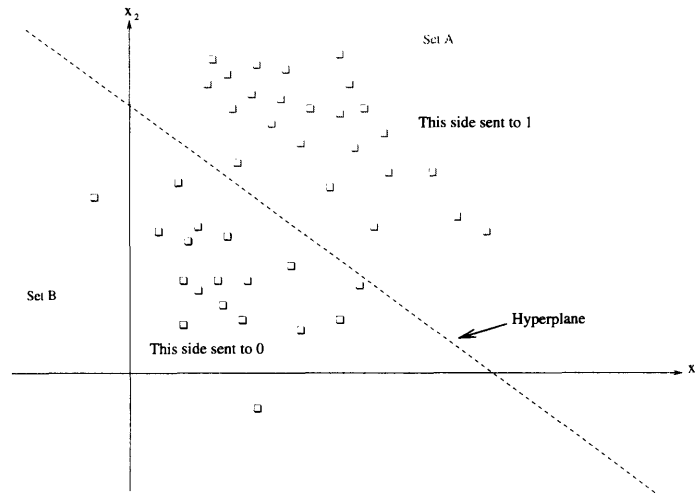


Figure 30: Hyperplane separating sets A and B .

In the more general case where A and B are not linearly separable more nodes must be used. For instance, the distribution in figure 31 can be separated by the node configuration shown in figure 32.

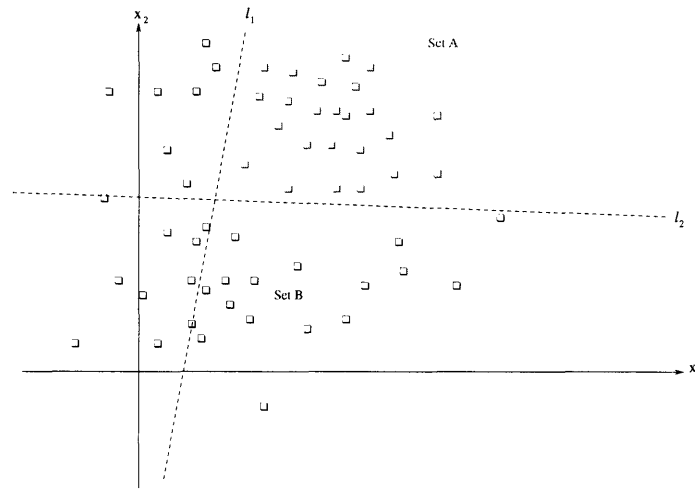


Figure 31: Graph of two sets A and B which are not linearly separable.

Here n_1 sends points to the right of l_1 to 1 and the rest to 0. n_2 sends points above l_2 to 1 and the rest to 0. n_0 returns 1 for $n_1 = 1$ and $n_2 = 1$ and returns 0 otherwise. It can be shown that all continuous functions whose domains are closed and bounded in \mathcal{R}^n can be approximated to any degree of accuracy by a neural network

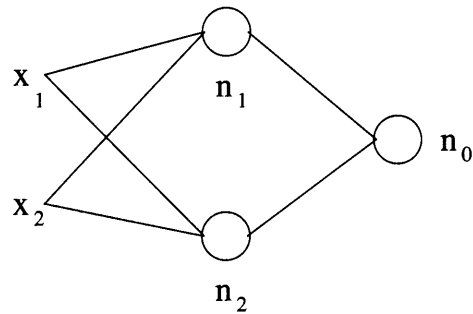


Figure 32: Node configuration which separates sets graphed in figure 31.

of one hidden layer with sigmoid or hyperbolic tangent activation functions[17]. For instance the sets graphed in figure 33 can be separated by a neural network with the node configuration shown in figure 34.

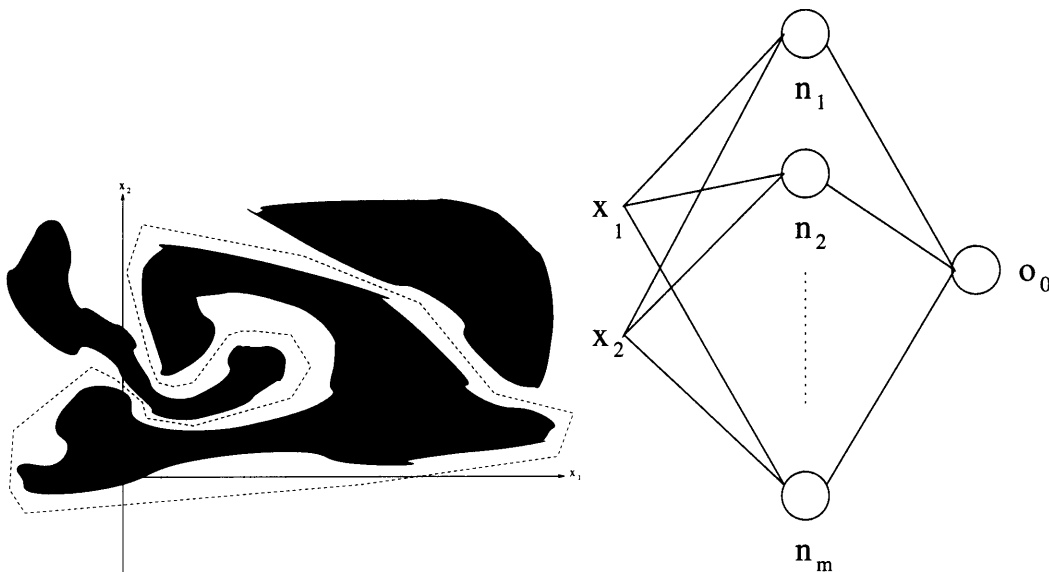


Figure 33: (Left) Two arbitrary sets which are separable by a neural net with one hidden layer.

Figure 34: (Right) A neural net with one hidden layer of m nodes which can be used to separate the two sets graphed to the left.

A.2 Training

A procedure called training is used to set the weights of the neural network. The neural network is trained using sets of data points for which the target output is known. Consider the case of the single layer neural network as diagrammed in figure 35.

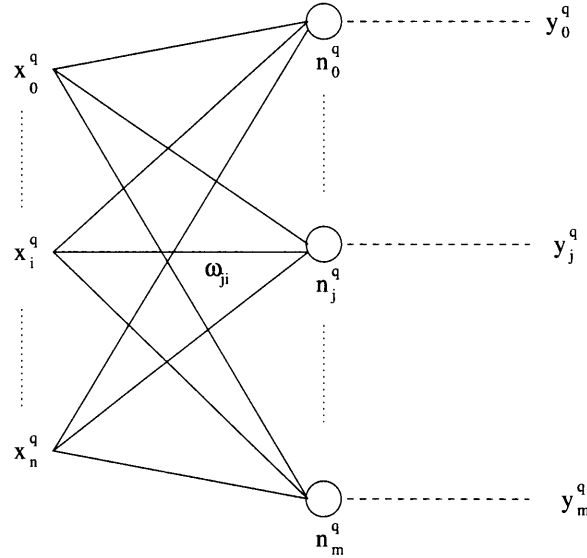


Figure 35: A neural network with $n+1$ inputs, $m+1$ nodes and $m+1$ outputs - one for each node.

Suppose there are N sets of $n+1$ data points $\{x_0^q \dots x_n^q\}$ and N sets of corresponding target outputs $\{y_0^q \dots y_m^q\}$ for the nodes, where the output of the j 'th node using the q 'th set of data points as inputs is n_j^q . The training procedure finds the values of the weights which bring the outputs of the nodes closest to the target outputs. This corresponds to finding the values of the ω_{ji} 's such that the sum squared error, $SSE = E = \sum_{q=1}^N E^q$, is minimized.

$$E^q = \frac{1}{2} \sum_{j=0}^m (n_j^q - y_j^q)^2 \quad (17)$$

Starting from a given set of ω_{ji} 's E is minimized by following its gradient with respect to the ω_{ji} 's in small steps downward. For each step the change in each weight ω_{ji} is

given by $\Delta\omega_{ji} = -\eta \frac{\partial E}{\partial \omega_{ji}}$, where η is the learning step size.

$$\frac{\partial E}{\partial \omega_{jk}} = \sum_{q=1}^N \frac{\partial E^q}{\partial \omega_{jk}} \quad (18)$$

$$\Delta\omega_{jk}^q = \frac{\partial E^q}{\partial \omega_{jk}} \quad (19)$$

By substituting

$$n_j^q = f \left(\sum_{i=0}^n \omega_{ji} x_i^q \right) \quad (20)$$

into (17) and differentiating it can be shown:

$$\Delta\omega_{jk}^q = \frac{\partial E^q}{\partial \omega_{jk}} = (n_j^q - y_j^q) f' \left(\sum_{i=0}^n \omega_{ji} x_i^q \right) x_k^q \quad (21)$$

Therefore:

$$\Delta\omega_{jk} = -\eta \sum_{q=1}^N \frac{\partial E^q}{\partial \omega_{jk}} = -\eta \sum_{q=1}^N \left((n_j^q - y_j^q) f' \left(\sum_{i=0}^n \omega_{ji} x_i^q \right) x_k^q \right) \quad (22)$$

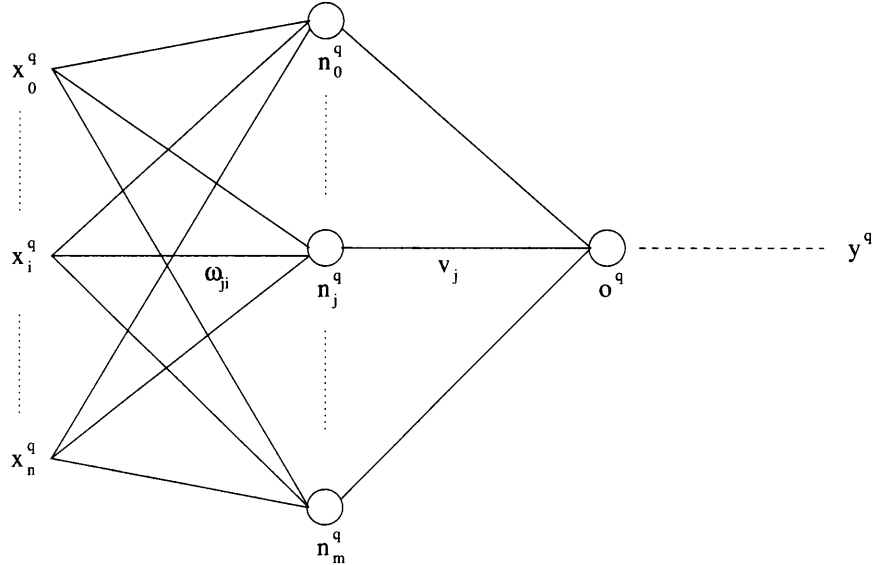


Figure 36: A neural network with $n+1$ inputs, a single layer of $m+1$ hidden nodes, and one output.

After each step a new cycle is begun, the gradient of E is recalculated, and a new step is made. These cycles are repeated until the minimum value of E is reached. For multiple layer neural networks, such as the ones used in BTagger, the training procedure is essentially the same except that the formulas for the gradient of E are more complicated. A neural network with the structure of a typical tagger neural network is diagrammed in figure 36.

In this notation the output of the j 'th hidden node when the neural network is run on the q 'th set of training data is $n_j^q = f(\sum_{i=0}^n \omega_{ji}x_i^q)$, the neural net output is $o^q = f(\sum_{l=0}^m \nu_l n_l^q)$, and $SSE = E = \sum_{q=1}^N E^q$ where $E^q = \frac{1}{2}(o^q - y^q)^2$. Only the target output for the last node y^q is known. Again the weights, ω_{ji} and ν_j are found by adjusting these weights by $\Delta\omega_{ij}$ and $\Delta\nu_j$ until E is minimized. The derivation of the formula for $\Delta\nu_j = -\eta \sum_{q=1}^N \frac{\partial E^q}{\partial \nu_j}$ is analogous to the derivation of the formula for $\Delta\omega_{ji}$. It can be shown that:

$$\Delta\nu_j = -\eta \sum_{q=1}^N \frac{\partial E^q}{\partial \nu_j} = -\eta \sum_{q=1}^N \left((o^q - y^q) f' \left(\sum_{l=0}^m \nu_l n_l^q \right) n_j^q \right) \quad (23)$$

$\Delta\omega_{ji} = \sum_{q=1}^N \Delta\omega_{ji}^q = -\eta \sum_{q=1}^N \frac{\partial E^q}{\partial \omega_{ji}}$ is derived as follows:

$$\frac{\partial E^q}{\partial \omega_{ji}} = \frac{\partial E^q}{\partial n_j^q} \frac{\partial n_j^q}{\partial \omega_{ji}} \quad (24)$$

Substituting for n_j^q it can be shown:

$$\frac{\partial n_j^q}{\partial \omega_{ji}} = f' \left(\sum_{l=0}^n \omega_{jl} x_l^q \right) x_i^q \quad (25)$$

$$\frac{\partial E^q}{\partial n_j^q} = \frac{\partial E^q}{\partial o^q} \frac{\partial o^q}{\partial n_j^q} \quad (26)$$

By substituting the formulas for o^q and E^q into equation (26) it can be shown:

$$\frac{\partial E^q}{\partial n_j^q} = (o^q - y^q) f' \left(\sum_{l=0}^m \nu_l n_l^q \right) \nu_j \quad (27)$$

Finally inserting (25) and (27) into (24):

$$\Delta\omega_{ji} = \sum_{q=1}^N \Delta\omega_{ji}^q = -\eta \sum_{q=1}^N \left((o^q - y^q) f' \left(\sum_{l=0}^m \nu_l n_l^q \right) \nu_j f' \left(\sum_{l=0}^n \omega_{jl} x_l^q \right) x_i^q \right) \quad (28)$$

Minimizing the SSE is analogous to minimizing the mean squared error $MSE = \frac{1}{N}E$.